Etudes biophysiques de l'interaction entre la protéine humaine TRBP et un précurseur de microARN oncogène

Matthieu Benoit

To cite this version:

Matthieu Benoit. Etudes biophysiques de l'interaction entre la protéine humaine TRBP et un précurseur de microARN oncogène. Autre. Université de Grenoble, 2013. Français. NNT : 2013GRENV015 . tel-01038628

HAL Id: tel-01038628

https://theses.hal.science/tel-01038628

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Biologie Structurale et Nanobiologie**

Arrêté ministériel : 7 août 2006

Présentée par

Matthieu Benoit

Thèse dirigée par **Jérôme Boisbouvier**
et codirigée par **Michael Plevin**

préparée au sein **du Groupe de Résonance Magnétique Nucléaire**
à l'Institut de Biologie Structurale Jean Pierre Ebel
et de **Ecole Doctorale Chimie et Sciences du Vivant**

Etudes biophysiques de l'interaction entre la protéine humaine TRBP et un précurseur de microARN oncogène

Thèse soutenue publiquement le **5 juillet 2013**,
devant le jury composé de :

Dr. Carlo Petosa
Institut de Biologie Structurale, Président
Dr. Bruno Kieffer
Ecole supérieure de biotechnologie de Strasbourg et Université de Strasbourg, Rapporteur
Dr. Carine Tisné
Université Paris Descartes, Rapporteur
Dr. Jérôme Boisbouvier
Institut de Biologie Structurale, Co-Directeur de thèse
Dr. Michael Plevin
Université de York, UK, Co-Directeur de thèse

There are always additional alternatives in an explanation of an observed phenomenon, corresponding to these unknown processes, both certain and inconceivable.

Contents

Acknowledgements

I performed this thesis in the NMR Group at the "Institut de Biologie Structurale Jean Pierre Ebel" (IBS) under the co-direction of Jerôme Boisbouvier and Michael Plevin. I really appreciated the wide range of NMR expertise and the open minded spirit of all the members of the NMR group which is the source of a very efficient cohesion. This culminated each week during the group seminars that were highly interactive and that I particularly appreciated. Having two supervisors was a rich situation in which I could benefit from the expertise of both of them. I especially want to thanks Jérôme to have accepted me in the NMR group and Michael that supervised closely the whole thesis as a main supervisor. I'm also grateful to my two PhD supervisors for their willing to have me participating to formations, seminars and congresses.

I also want to thank all the members of the NMR group and especially Lionel Imbert that undertook the production of the RNAs. This collaboration was a key point for the success of the protein/RNA interaction experiments. I also really enjoyed sharing an office with him during three years. I want to thanks Isabel Ayala and Ombeline Pessey for their assistance in the use of the protein production and purification devices, and for the very good atmosphere. I also want to thanks Julien Perard for the assistance in the recording of MALLS experiments and Andres Palencia for his practical assistance in the design of ITC experiments. Eventually I want to acknowledge Christine Ebel & Aline LeRoy that spend time to show me how to perform svAUC experiments and to use programs to analyze svAUC data. I also thank Marjolaine Noirclerc-Savoye and Benoit Gallet (RoBioMol at the Intitut de Biologie Structurale, Grenoble) for the initial construction and testing of the TRBP constructs.

I especially want to acknowledge the CEA for a 3 year fellowship from the International PhD program (IRTELIS) and the French ARC that gave me the opportunity to go on my thesis work during the last 6 months of my thesis. The conclusions from the ARC project are under construction at the time of writing and will be presented during the thesis defense.

I want to thank all the members of the Jury with whom I had a very interesting question session.

I'm also gratefull to all the developpers of the GNU R and GNU Emacs programs that became indispensable tools for me to perform reproductible research.

I also want to acknowledge all my family and especially my parents and my sister who supported me during my PhD, and all the friends with whom I shared my passion of field ornithol-

ogy during this period.

Abstract

MicroRNAs (miRNAs) are a class of small non-coding RNAs that regulate gene expression through RNA interference (RNAi). Human miRNAs are generated via a series of enzymatic processing steps. In particular, in the cytoplasm, the precursor miRNA (pre-miRNA) is recognized and cleaved by a complex containing the RNase III enzyme Dicer and several non-catalytic accessory proteins. HIV TAR element-binding protein (TRBP) is a constituent of the Dicer complex, which augments complex stability, has effects on the cleavage kinetics and on the cleavage site and potentially functions in substrate recognition and product transfer to the RNA-induced silencing complex (RISC). TRBP is composed of three double stranded RNA binding domains (dsRBDs). The RNA binding region of TRBP is composed of the first two dsRBDs and an uncharacterized interdomain region. The present study reports the *in vitro* biophysical characterization of the RNA binding region of TRBP in the apo state and in the RNA bound state with the two successive cytoplasmic precursors of the oncogenic human microRNA miR-155, the hairpin pre-miR-155 and the related Dicer product miR-155/miR-155* duplex. The study shows that the RNA binding region of TRBP is monomeric and comprises two independent double-stranded RNA-binding domains connected by a 60 residues flexible linker. The first dsRBD, uncharacterized previously in solution, undergoes a full folding/unfolding equilibrium in a wide range of physico-chemical conditions. The two first dsRBDs of TRBP can interact with one microRNA precursor and two RNA binding regions can interact with one precursor molecule. The RNA-binding region of TRBP interacts with both pre-miR-155 and miR-155/miR-155* duplex with similar affinities. In the complex with one RNA binding region of TRBP bound to either pre-miR-155 or miR-155/miR-155* duplex, no evidence of contact between the two dsRBDs were observed and the protein interacts with both precursors via the same protein binding surface. The data presented here suggest that the RNA binding region of TRBP can play a role before and after processing of pre-miRNAs by Dicer, including in the RISC loading complex.

Motivations

General motivations Various aspects have guided the choice of the system studied in the present work. First, at the time of the subject definition, microRNAs (miRNAs) represented a recent field discovered a decade ago that revolutionized the conception of gene regulation and the design of *in vitro* gene expression experiments. MiRNAs are small RNAs that regulate post transciptionaly gene expression of most cell mechanisms in a wide variety of multicellular eukaryotes. The clear involvement of miRNAs in cancers has been proved for a wide variety of neoplasias and as a result miRNAs have started to gain interest for potential applications in oncology. As a consequence of these interests over the past decade, there has been a drive to understand the biology of the regulation by miRNAs and the structural biology of miRNA biogenesis. For the latter goal, X-ray crystallography, electron microscopy and NMR spectroscopy have been used to get insight into some complexes involved in the biogenesis. However most of the structural data have been obtained on unbound parts of the proteins. There was a general lack of structure and dynamics studies of complexes allowing to describe the location of the components, to determine how one component interacts with others and to understand how these complexes work in miRNA biogenesis.

In particular, one key protein called TRBP is part of all the cytoplasmic complexes involved in miRNA biogenesis and this has attracted our attention. It is thought to be involved in the recognition of miRNA precursors, which seems a challenge since there are more than 1000 different precursors in humans. Many biochemical studies have sought to explore the mechanism for RNA recognition by TRBP or examine how TRBP associates with other proteins, in particular with the enzyme Dicer. Structural studies of TRBP have elucidated the isolated 3D structures of the first two double stranded RNA binding domains (dsRBDs) as well as the second dsRBD in complex with a short siRNA duplex. Negative stain electron microscopy studies of miRNA biosynthesis complexes containing TRBP have also been reported, however the level of resolution was not yet sufficient to accurately describe the localization of TRBP in these complexes or to determine how it interacts with other components. Thus despite these efforts, the exact role and mechanism of TRBP in human miRNA precursor recognition and processing remained unclear. In particular the way the multiple RNA binding domains could interact with the miRNA precursors was not known, neither the stoichiometry of these complexes. Some of the main questions were: Are all the domains involved? Is there one precursor binding each domain? Where do the domains bind on the miRNA precursors? etc. Moreover, some properties of the protein within the biogenesis pathway could not be explained with the limited structural data available.

Additional properties of TRBP made it an interesting protein to study. It contains double stranded RNA binding domains, a fold which is widely present in the proteins involved in miRNA biogenesis in all the organisms possessing this biological pathway. Various studies have demonstrated an important variability of the behavior and roles of these domains. Despite it was first supposed that these domains were all equivalent, it is now clear that each one has a special set of properties. As more and more dsRBDs are studied, a pleiotropic character of the behavior and dynamics of these domains is emerging. Thus, working on TRBP was also attractive because of the possibility to discover new behaviors and to expand the knowledge on these variable domains.

Personal motivations In addition to the general motivations that led me to choice this thesis theme, I had a strong methodological interest. I had previous experience in proteomics, an active field that can provide a wide, multidimensional view of the states of cells. Despite lots of information and discoveries can result from this field, I've been frustrated that it was often impossible to explain mechanistically what was going on without high resolution, site-specific information on the macromolecules. This is the main goal of structural biology. One of the most used technique able to provide high resolution structure of macromolecules - the study of x-ray diffraction by molecules crystals - is sometime inducing states in the crystal that are irrelevant in solution. Taking into account in particular that many proteins involved in miRNA biogenesis, and in particular TRBP, were predicted from their sequence to be highly flexible, it was necessary to perform the studies in solution. Liquid state NMR spectroscopy appeared as a method of choice for tackling human miRNA biogenesis complexes in solution with high precision information. However, this technique couldn't easily answer all the questions on this complex. In order to get relevant in depth information on the system, it was required to use multiple complementary and compatible in solution biophysical techniques.

Thesis contents The present manuscript classically starts (part I) by a review of the available data of the biological system studied and analogous systems in the small silencing RNA context in multicellular organisms. The second part is devoted to an introduction of the biophysical techniques used with the aim to explicit the information required for an appropriate understanding of the experimental work performed. The practical implementation of these techniques adopted in this work in also included in this part. The next part (part III) presents the production strategy of the miRNA binding region of TRBP and of the cytoplasmic precursors of the oncogenic human miRNA miR-155. The biophysical characterization of the isolated protein and RNA macromolecules are then presented (part IV), before the macroscopic and site specific studies of the RNA binding region of TRBP with miR-155 precursors.

Part I

Polymorphism and pleiotropy of dsRBDs in mi/siRNA biogenesis

This part aims to present the context of the work performed in this thesis by critically reviewing the literature and analyzing some structural data related to the role of double stranded RNA Binding Domains (dsRBD) in the biogenesis of two classes of small silencing RNAs, microRNAs (miRNAs) and short interfering RNAs (siRNAs) in multicellulars. The presentation is split in four chapters, which will go from the most general considerations to the more detailed ones on dsRBDs :

- **Chapter 1**: General presentation of miRNAs and small silencing RNAs. The chapter sets the rich biological context by focusing on the microRNAs and siRNAs as post-transcriptional regulators in multicellulars.

- **Chapter 2**: Roles of dsRBD containing proteins (dsRBP) in miRNA and siRNA biogenesis. This is the biological functional context of the work performed in this thesis. The requirement of multiple proteins containing dsRBDs in mi/siRNA biogenesis is underlined. The chapter presents the very diverse roles of each dsRBP protein along the mi/siRNA pathways, including the network of interactions, but without mechanistic details. This chapter will underline the abundance of dsRBDs at the different stage in the biogenesis of miRNAs and also the variety of miRNA pathways (i.e. diverse substrates to be recognized) in different model organisms. This chapter is required for an appropriate biological understanding of chapters 3 and especially 4.

- **Chapter 3**: Sequence and fold comparison of the dsRBDs involved in miRNA and siRNA biogenesis. The second chapter is devoted to a comparison of the chemical and structural features of dsRBDs involved in microRNA biogenesis, including original statistical analyzes on dsRBDs structures underlying the polymorphism of dsRBDs in mi/siRNA biogenesis. A partition of dsRBDs fold in their apo state into 4 classes in proposed and discussed.

- **Chapter 4**: Comparison of specific roles of dsRBDs in dsRBP mi/siRNA biogenesis proteins. The comparison is made transversally for different organisms and pathways to emphasize the pleiotropy of dsRBDs in mi/siRNA biogenesis. Such a review on the role of dsRBDs in multicellular miRNA and siRNA biogenesis is original and was not performed before. It suggests multiple research directions.

To avoid repetition and to better understand the biological context of this thesis, the data specific to the protein TRBP studied in this thesis is spread along chapters 2, 3 and 4.

Chapter 1

MicroRNAs and short interfering RNA silencing in multicellulars

This chapter focuses on the main types of microRNA and siRNA biogenesis pathways in multicellulars, underlying the similarities of the different pathways and in particular the abundance of dsRBDs related to these pathways. Prokaryotes also have an RNAi based mechanism but it is only analogous (not homologous) to the one of eukaryotes (245). The present bibliographical review only focuses on the RNAi mechanisms in eukaryotes.

1.1 History and definitions in the small regulatory RNAs world

Discovery of a gene regulation mechanism based of RNA interference in multicellulars. Before the start of the 1990s, the known mechanisms of gene regulation could be partitioned into three subsets: pretranscription regulation of the chromatin state, transcription regulation and translation regulation. After various experiments in the 1990s, when some groups managed to inhibit genes by dsRNAs, in 1993 a group found in *Caenorhabditis elegans* (*C. elegans*) that the gene lin-4 codes for a small RNA that inhibit the expression of the product of the gene lin-14 via an antisense RNA/RNA interaction (164). From 2000, progressive identification of small regulatory RNAs present in different plant and animal species proved the existence of a relevant general phenomenon. These small RNAs of about 20 to 30 nucleotides mediate what is called RNA interference or RNA silencing. This phenomenon is the inhibition of the expression of a specific messenger RNA, either by degradation of the RNA or by the inhibition of its translation, in which a small-RNA/mRNA base pairing is involved. The two terms silencing and interference coexists because an historical distinction that appeared to be irrelevant. The small regulatory RNA will be generically called small silencing RNA from now, an explicit expression already employed in the literature and with the advantage that it does not match with the ones devoted to previously recognized subgroups

Building a partition of the plethora of small silencing RNAs Deep sequencing studies in many eukaryotes have led to the discovery of a plethora of small silencing RNA. A tendency to name apparently original ones have led to a profusion of names for subgroups of small silencing

RNAs. The main groups are called microRNA (miRNAs), small interfering RNA (siRNA) and piwiRNAs (piRNAs). However, even with this simple classification, the definition of the first two classes are not consistent between different works and appear to strongly depends on particularities of the model organisms that is the subject of the article giving the definition. For instance the partial base pairing complementarity of mammalian microRNAs with their messenger RNA (mRNA) target is often used as a definition in mammals to distinguish them from siRNA with a perfect base pairing complementarity. However in mammals a few so called miRNAs are recognized to fully base pair with their target and in plants, miRNA base paring is almost always perfect (59). As a result, it seems important to assess at one stage whether the useful definition from the past or only valid in one species have still a value in light of the whole set of discoveries. Many variables are inconsistently used to define the different types of small silencing RNAs in the literature: their global origin (endogenous or exogenous), the structure of the precursor (duplex, hairpin...), the subset of protein required for their biogenesis, the length of the small silencing RNA and the extension of base pairing. The three main recognized classes of small silencing RNAs have similar length between 19 and 25 for miRNAs, 20 to 25 for siRNAs, and 26 to 31 for piRNAs and all trigger an mRNA silencing (degradation or translation inhibition) mediated by a complex containing a protein of the Ago type. It could appear tempting to unify all these small regulatory RNA in a larger group. However, these differences in biogenesis seem to have functional meaning. In mammals, miRNAs are known to be post transcriptional regulators of gene expression. Endo-siRNA pathway in mammals seems required for oocyte maturation and preimplantation and meanwhile miRNA function is repressed during these embryonic development steps (253). Another example, in plants the siRNA pathway is an antiviral mechanism for viruses with a dsRNA produced during the viral reproduction cycle. However, the biological function of miRNAs and siRNAs are sufficiently different to distinguish them on this basis (6). A classification using the proteins involved in the biogenesis has already been recognized as a good way to define classes of small silencing RNAs (99). Importantly it enables to define homogeneous groups of small RNA valid whatever the species among the ones under study.

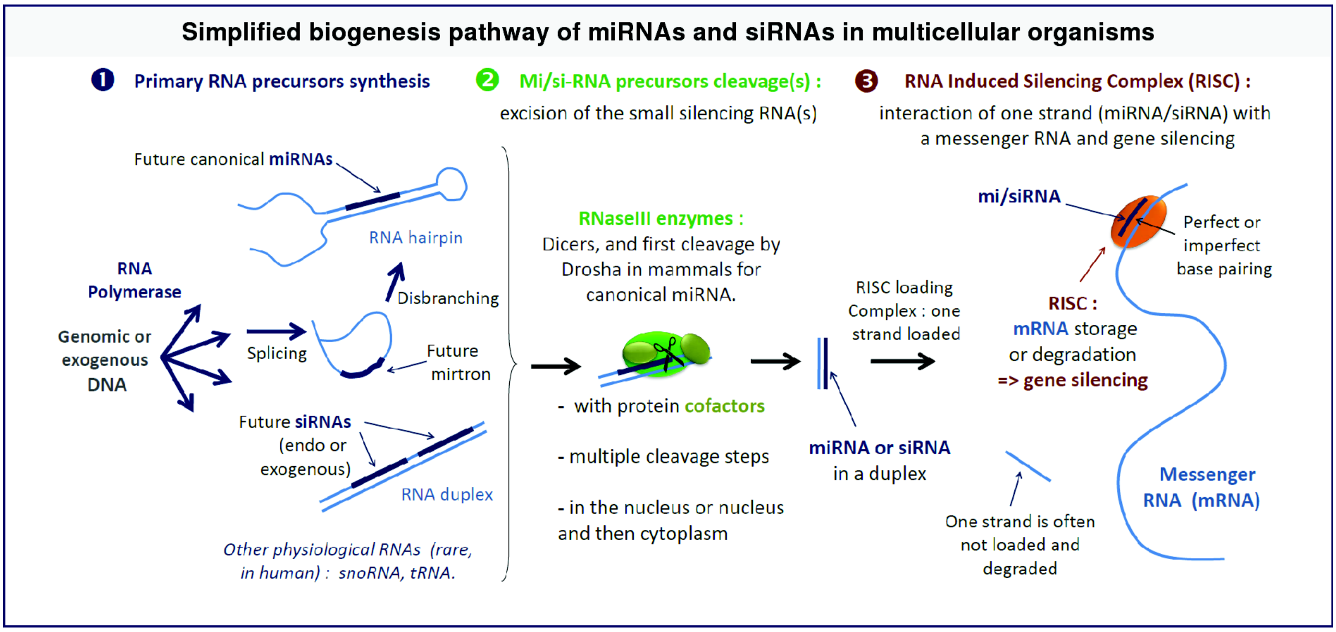


Figure 1.1: Simplified pathway illustrating the biogenesis of miRNAs and siRNAs in multicellular organisms in three main steps.

Definition of the main classes of small silencing RNAs Taking into account the previous paragraph, I propose the following definitions for 6 recognized classes of small silencing RNAs. A general simplified pathway of the biogenesis of siRNAs and miRNAs is presented in **fig. 1.1**. The following definition are not definitive and these groups are a way to discriminate the rather blur ensembles of small silencing RNAs.

- canonical microRNAs: *genome encoded* single-stranded RNA molecules, produced in eukaryotic cells from hairpin precursors generated from a primary RNA transcript containing at least one local hairpin structure, by two cleavages catalyzed by endoribonuclease III enzymes (6; 297). *Note* The cleavage are catalyzed by Drosha and/or Dicers and exceptionally without Dicer replaced by the Ago protein (e.g. hsa-miR-451: Drosha and Ago2 dependent for its biogenesis (38; 291; 293)).

- mirtron microRNAs: endogenous single-stranded RNA molecules, produced in eukaryotic cells from spliced-out of introns with a hairpin potential and a debranching enzyme, and then cleaved by a Dicer enzyme (279).

- miRNA derived from functional RNA requiring Dicers for their processing: small nucleolar RNA (snoRNA) and tranfer RNA (tRNA)) *Note* Small nucleolar are RNAs of 70-200 nucleotides in length that participates in nuclear ribonucleoprotein complexes where they serve as antisense guide, like miRNAs. SnoRNAs are abundant, well conserved and participate in nucleotide modification on some ribosomal RNA (189). The dicer requirement for miRNA derived from snoRNA in humans has been shown (71) and in mice (9). These miRNAs (198) are mentioned here for the sake of exhaustivity ; the extensions of these classes of miRNAs are unclear. The regular discovery of original cases of small silencing RNAs suggests that it is likely that other classes of small silencing RNA exists.

- endo-siRNA: endogenous short/small interfering RNA (siRNA) originating from long endogenous duplexes and that require endoribonucleases III for their processing and require Dicer enzymes.

- exo-siRNA: exogenous version of siRNA. SiRNAs originally referred to exogenous RNA, widely used to silence genes experimentally *in vivo* and *in vitro*. *Note*: The sources can be either viral or biotechnological for research or medical goals and can correspond to short hairpin RNAs.

- piwiRNA: small RNA generated by a Dicer and Drosha independent pathway and associated to piwi class of Argonautes proteins (143). *Note* They are mainly derived from transposons and repeated elements. In *Drosophila melanogaster*, Zebrafish and mouse, they are involved in the germline development, germline DNA intergrity and silencing of selfish DNA elements, reviewed in (143)

Focus of miRNAs and siRNAs Among the 6 previously defined classes of small silencing RNAs, piRNA represents the most distinct (and abundant) class, with very different proteins involved in the biogenesis, a lack of piRNA sequence conservation in contrary to miRNA and siRNA and distinct biological effects. The experimental work in this thesis has been performed with human TRBP and miRNA precursors, so present development will not include piRNA but will focuses on both miRNA and siRNAs. Both miRNA and siRNA pathways appear to be interconnected. The idea is that it is only by taking into account the large diversity (families and members of these families) of small silencing RNA undergoing biogenesis with shared complexes that the role of the member of these complexes (including the dsRBDs) will be adequately understood. Indeed TRBP protein is involved in both miRNA and siRNA biogenesis. Another interest of reviewing the literature on siRNAs is that many experimental biophysical data, especially in mammals were performed with siRNAs or siRNA like molecules rather than miRNA. This is because of the high interest in exogenous siRNA as a research and medical tool. Additionally, a review not centered on TRBP but covering the role of the dsRBD containing protein in multicellular miRNA and siRNA biogenesis enables to put the known properties of TRBP in a broader evolutionary context.

Nomenclature of miRNA and siRNA MicroRNA names (6) are formed by the concatenation of :

- a species prefix that identify the organism (e.g. "hsa-" for human)

- the word "mir-" for miRNA genes, "miR-" for mature miRNA in mammals and "MIR-" in plants

- a number assigned to a miRNA sequence whatever the organism. The same number can also be assigned to very similar miRNAs.

- a possible letter to distinguish very similar miRNAs

- sometimes the words "-5p" or "-3p" to precise which strands (5′ or 3′) of the precursor stem that contains the miRNA. Recent data indicate that for many cases both strands of the miRNA duplex can be loaded, but usually with a clear strand bias, one strand being more loaded on a messenger RNA than the other one. The notation using a star (*) after the name of the miRNA indicate a strand much less loaded or not loaded on a messenger RNA. Example: hsa-miR-155-5p, the duplex hsa-miR-155/hsa-miR-155*

SiRNAs are commonly named using their target gene name, the word "-siRNA-" and a number to distinguish them. For instance : p53-siRNA-2.

1.2 Properties and biological functions of miRNAs and siRNAs

The definitions of the previous classes of small silencing RNAs are used here to present the main properties of miRNAs and siRNAs although many do not strictly differs between the classes. They are presented in parallel on common themes to facilitate the comparison. The biogenesis of the different types of miRNA and siRNA are presented in the next part.

1.2.1 Phylogenetic repartition of miRNAs and siRNAs

miRNA An RNAi machinery seems to have appeared before the emergence of eukaryotes, from archeal, bacterial and phage proteins implicated in DNA repair and RNA processing (245). The key proteins involved in miRNA biogenesis are conserved in plants and metazoans. On the contrary, the homology between miRNA exists in both plant and metazoans but is not shared between these two clades, indicating a separate evolution of miRNA in these two groups from a common ancestor (245). RNAi may have been originally selected as an antiviral mechanism against viruses with a double stranded RNA formed during their replication cycle. miRNAs are present in plants including some unicellulars (e.g. unicellular green alga, *Chlamydomonas reinhardtii*), in metazoans and in the virions that can infect them. Existence of the mirtron microRNA pathway has been shown in Drosophila melanogaster, in *C. elegans* (37), mouse and humans (152). The fungi possess proteins of the RNAi machinery but have not functional miRNAs.

endo-siRNA This recently found class of small RNA seems ubiquitous in eukaryotes, has been found in plants, *C. elegans*, *Drosophila melanogaster* (somatic and germ cells) and in mammals (99) including mouse oocyte, male germ cells (253) and mammalian brain (251) where is seems to have a role in neuronal gene expression.

1.2.2 Number of known miRNAs and siRNAs in humans

A general database exists for non coding RNAs (http://biobases.ibch.poznan.pl/ncRNA). Identified microRNAs are stored and annotated in the miRBase (http://microrna.sanger.ac.uk/). The number of entries in the database at the time of writing is about 22000, some of them being conserved in all the organisms. About 1000 miRNAs are known in human and many of them have variants. About 240 mirtrons miRNAs are predicted in mouse and humans (152). The number of the recently discovered endo-siRNAs in humans is still nuclear

1.2.3 Chemical structure of precursors and mature miRNAs and siRNAs

miRNAs and siRNAs are polymers of Uridine, Guanosine, Adenosine, Cytosine and Inosine. The ends are 5′-phosphorylated and 2′-3′-hydroxylated apart in plants and some metazoans where miRNAs are 2′OH methylated. A common property of these RNAs is their size, about 22 nucleotides (nt) +/- 3 nt. microRNAs in mammals have usually bulges and mismatches in their structure. From here, most microRNA illustrations will be given with hsa-miR-155-5p which will be written miR-155 and its complementary strand, miR-155*. MiR-155 is one of the best known miRNA, recently reviewed in (70). It is produced from the gene MIR155HG (entrez gene id: 114614). Both the 5p and 3p(*) strands of the microRNA 155 can be loaded on a messenger RNA and lead to mRNA inhibition, but the 5p version, oncogenic, is the most known. The example of hsa-miR-155-5p is given in **fig. 1.2** at different steps of the miRNA pathway and with one of its target, the messenger RNA of the tumor protein 53-induced nuclear protein 1 (TP53INP1).

1.2.4 Targets and effects of miRNAs and siRNAs

MiRNAs and siRNAs silence genes by mRNA degradation or inhibition of translation (253), the exact mechanism depends on the organism as well as the extent of base pairing with the mRNA. MiRNAs mainly target the 3′ untranslated region of mRNAs. They act as post-transcriptional regulators of gene expression. All known plant microRNAs have a perfect base pairing with their target, leading to the cleavage of the mRNA. On the contrary, most animal miRNAs have an imperfect base pairing with their target mRNAs like miR-155. The miRNA/mRNA interaction triggers mRNA silencing with association with protein factors of the RISC complex that promote an increase of mRNA decay or repress the translation of the mRNA (81). The non perfect base paring can be restricted to only 7 base pairs (21). The miRNA targets appear determined by the so-called seed region which is located between nucleotides 2 and 8 the mature miRNA. MicroRNAs can target very different processes in the cells ranging from development, cell differentiation, cell division and apoptosis (27). However, it is still unclear what is the amount of genes regulated by miRNAs. Bioinformatics studies predict that about 60% of eukaryotic genes could at least partially match in humans (88) but this does not ensure that the gene is significantly regulated by the miRNA, the ratio between miRNA and mRNA being an important factor to take into account.

SiRNAs target viruses, mobile elements and can regulate chromatin assembly in some organisms. The silencing mechanism is very similar than the miRNA one. Their artificial exogenous counterparts are artificially designed to knock down gene function *in vivo* and *in vivo*.

A. **pri-hsa-miR-155, before Drosha cleavage**

```
5' *******************           c   a        u-   cc
                      uuaaugcuaau gug uagggguu  uug  \
                    acaauuacgauua-uac-auccucag  aac  /
3' *****************                          uc   cu
```

B. **pre-hsa-miR-155, before Dicer cleavage**

```
                                 c   a        u-   cc
                  5'P-uuaaugcuaau gug uagggguu  uug  \
               3'OH-acaauuacgauua-uac-auccucag  aac  /
                                              uc   cu
```

C. **hsa-miR-155-5p / hsa-miR-155-3p duplex**

```
                                 c   a
                  5'P-uuaaugcuaau gug uaggggu-3'OH            hsa-miR-155-5p
               3'OH-acaauuacgauua-uac-auccuc-5'P              hsa-miR-155-3p(*)
```

D. **hsa-miR-155-5p with TP53INP1 mRNA**

```
               5'P-uuaaugcuaauc---guga-uaggggu-3'OH              hsa-miR-155-5p
                    ||||||||||    |||  ||
3'***cguugauaguggacaauuacgauuacaaucacacauuuuuacauaaguaguaua***5' mRNA TP53INP1
```

Figure 1.2: Predicted secondary structure of the human oncogenic miRNA hsa-miR-155-5p precursors at different steps of the biogenesis using information from miRbase. **A.** Secondary structure of the primary miRNA 155. Because of the long length of the miRNA primary transcript and TP53INP1 mRNA, most of their sequence is represented by a series of starts (***) of arbitrary length. **B.** Secondary structure of the pre-miRNA 155 after cleavage by Drosha/DGCR8 (microprocessor) complex. **C.** Secondary structure of the miRNA/miRNA* duplex of miRNA 155. **D.** Secondary structure of the predicted interaction sites between hsa-miR-155-5p and one of its target messenger RNA TP53INP1 in the 3′ untranslated region (100) at the RISC level in the miRNA pathway. The interaction shows the so called seed sequence at nucleotides 2 to 8, which is the main base pairing region, as well as additional interactions. The base pairing is imperfect as usual in mammalian miRNAs.

1.2.5 Involvement of miRNAs and siRNAs in diseases

Because of the wide range of processes regulated by miRNAs, some are involved in human diseases (262) and especially in cancers, reviewed in (94; 205; 165). Some miRNAs have oncogene or tumor suppressor. Hundreds of miRNAs are dis-regulated in cancers cells. MiRNAs have been found to be involved in all the steps of tumorigenesis. Interestingly, tumor prognosis using miRNAs expression profile is more reliable than prognosis based on messenger RNAs. Involvement of non canonical miRNAs and endogenous siRNAs in diseases is yet unclear but expected according to (287). The canonical human microRNA miR-155 (5p) has causal effects in various cancers and diseases (70) including acute lymphoblastic leukemia, diffuse large B-cell lymphoma, breast cancer, pancreatic cancer, hepatocellular carcinoma and hypertension. For instance, hsa-miRNA-155-5p silences the pro-apoptotic Tumor Protein 53 Induced Nuclear Protein 1 (TP53INP1) **fig. 1.2**. Where hsa-miR-155-5p is overexpressed, this microRNA contributes to the lack of apoptosis in cells, one of the phenotype contributing to cell malignancy. Mutations in the machinery responsible of mi/siRNA biogenesis have also been reported to be linked with cancers. For instance in human Dicer, there have been more than forty human heterozygous germline mutations reported worldwide (85), most of them related with cancers. A mutant of the protein TRBP lacking its last domain has also been reported in cancers (193). In addition, the antibiotic enoxacin has been shown to be a cancer specific growth inhibitor, an effect that was proposed to be related with the interaction of this drug with TRBP (192). The truncated oncogenic mutant of TRBP does not interacts with this drug (192).

Chapter 2

Diversity of the functions of dsRBPs in mi/siRNA biogenesis

From the discovery of microRNA, biogenesis pathways were studied in different model organisms. The key protein members and the order of their contributions are rather well understood. Some proteins involved in miRNA biogenesis like the human TRBP were actually discovered before the microRNAs (97). A large diversity of miRNA and siRNA pathways exists in different organisms. For the purpose of this review, a selection of organisms had to be made. Model organism belonging to different multicellular clades with abundant data available on dsRBD containing Proteins (dsRBPs) including functional roles, interactions and structures were selected: the vertebrate *Homo sapiens*, the worm *Caenorhabditis elegans*, the insect *Drosophila melanogaster* and the plant *Arabidopsis thaliana*. A few relevant data are also included from various rodents and the Frog *Xenopus Leavis* when useful.

2.1 dsRBDs are a common feature in proteins involved in small RNA biogenesis

There is always at least one RNaseIII enzyme involved in mi/siRNA biogenesis, an RNA interference silencing complex with an Ago protein, and some dsRBD containing proteins (dsRBP) involved along the biogenesis. When multiple organisms and small silencing RNAs are compared, the most common RNase III enzyme in miRNA and siRNA biogenesis appears to be Dicer. Taking into account the recently discovered non canonical miRNA pathways, I propose here a split of miRNA and miRNA biogenesis pathways in multicellular into three main phases, developed in the next paragraphs: primary precursor production, generation of a short dsRNA from the precursor, and small RNA loading on a/some messenger RNA(s). These steps are the ones introduced in **fig. 1.1**.

Representative examples of dsRBD containing proteins involved in miRNA and siRNA biogenesis are presented for the model organism species covered here in **fig. 2.1**. It shows the primary domain organization of dsRBP proteins, underlying the abundance of the dsRBDs.

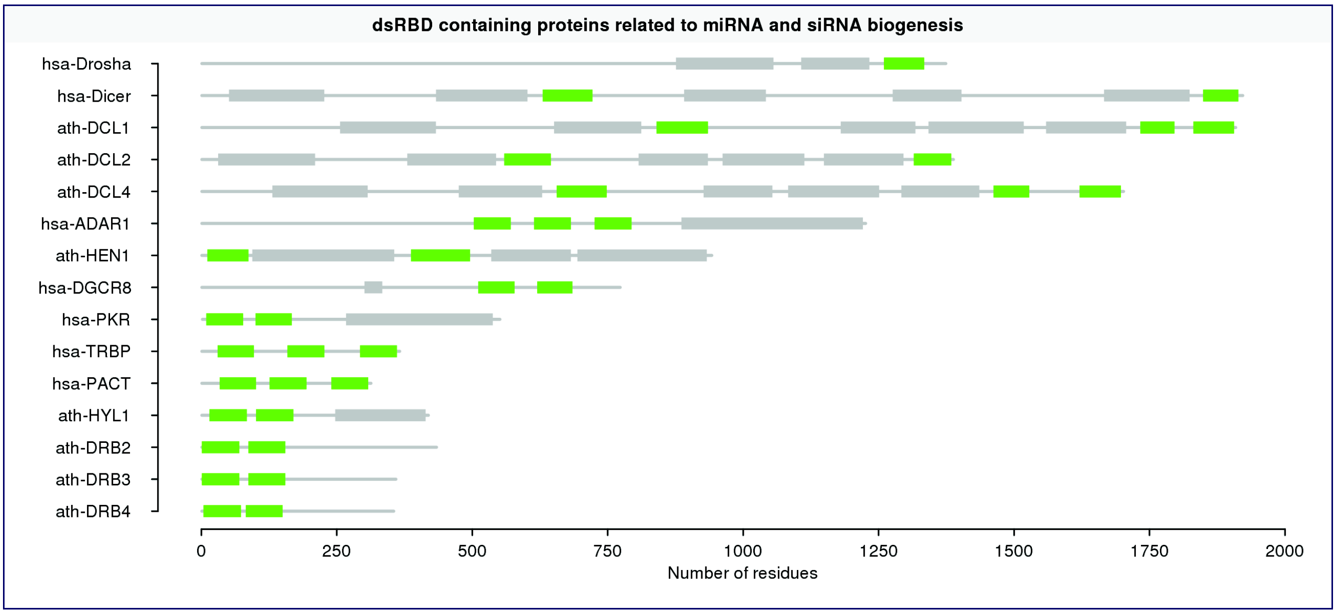


Figure 2.1: Cartoon view of members of the main classes of dsRBD containing proteins involved in mi/siRNA biogenesis in *Homo sapiens* and *Arabidopsis thaliana*. The relative lengths of the domains are proportional to their actual sequence length, obtained from the Uniprot database. The domain organization of *Arabidopsis thaliana* HEN1 was obtained from the available 3D structure of the protein (256). The axis below the figure represents the length in residues units.

The next section develops the main features in common in the miRNA and siRNA pathways of the selected organisms along the three main phases of the biogenesis. It might not be general rules as huge diversity exists in multicellulars ; the words metazoans and plants are used for convenience with this in mind. Then the specificities of the pathways of the selected model organisms are presented successively. This aim at introducing the multiple dsRBD containing proteins in their biological context, before a focus in the next chapter of their detailed roles in the biogenesis of small silencing RNAs where the presentation will no longer follows phylogenetic partitions.

2.1.1 Generation of mi/siRNA primary precursors and ADAR dsRBDs

Primary precursor synthesis The primary RNA (pri-miRNA or pri-siRNA) can be considered the first RNA structure that contains the dsRNA region containing the small silencing RNA that will be loaded by the RISC complex. Its nature is small RNA dependent :

- Canonical pri-miRNAs are generated from a defined locus by the RNA polymerase II (or sometimes III). The primary transcript of miRNAs can contain multiple hairpins and each of them can give rise to a different microRNA.

- the primary transcript of mirtrons miRNA corresponds to some gene introns in mRNA generated by a RNA polymerase and a debranching enzyme.

- for miRNA derived from other cellular RNAs (snoRNA and tRNA), the latter RNA are the primary transcripts.

- primary transcripts of endo-siRNAs are still poorly known.
- dsRNA part of viral replication cycle is often a primary transcript of siRNA in plants.

Metazoan pri-miRNA edition by the dsRBD containing ADAR enzymes In metazoans, the primary transcript generation phase can also include an RNA edition step (base editing, methylation), reviewed in (209). Metazoans double stranded RNA specific adenosine desaminases (ADAR) catalyze the conversion of adenosine (A) to inosine (I) by a deamination of double stranded RNAs. ADAR proteins have been studied in various metazoans (human, Rats, Frogs, etc). In human, about 16% of primary microRNAs are edited in the brain (136). Adenosine to inosine editing consequences are diverse: they can range from an inhibition of the processing of some pri-miRNA by Drosha (208), a diminution of RNAi to a change of the seed sequence of some miRNA as inosine preferentially interacts with cystidine that can be sufficient to change the miRNA target without interfering with the maturation process. The edition can be tissue dependent. In mouse, miR-376-5p edition changes its target from threonine and tyrosine kinase to different target including phosphoribosyl pyrophosphate synthetase 1, a change confirmed *in vivo* (138).

ADAR protein weight is about 120 to 150 kDa. The topology from the N-terminus to the C-terminus comprise typically two DRADA repeats, three double stranded RNA binding domains and an adenosine to inosine editase domain. As an example, the domain organization of ADAR1 of human is given in **fig. 2.1** .

2.1.2 pri-mi/siRNA cleavage to small dsRNA duplex involves many dsRBDs

The primary RNA precursor is cleaved by RNase III enzymes, often multiple times. These enzymes belong to the family Drosha (i.e. RNase III enzymes of class 2) and Dicers (i.e. RNase III enzymes of class 3), possessing at least one dsRBD, usually more. The cleavages can be spatially separated, like miRNA in metazoans where one cleavage occurs in the nucleus and the second in the cytoplasm after a nuclear export of the RNA. In such case, the intermediate RNA between the two cleavage reactions is called a "precursor" miRNA (pre-miRNA). The result of this phase is a double stranded RNA (dsRNA) of about 20 (+/- 3) nucleotides in length. Drosha and Dicers have systematically a set of dsRBD containing cofactors, often dispensable, that have multiple dsRBDs (2 to 3). Very few miRNAs have been reported to escape part of this phase, like hsa-miR-451 (38; 291; 293) integrated in the next phase as a hairpin and not as a dsRNA duplex.

Drosha In metazoans, the production of canonical pri-miRNAs and their processing to a precursor miRNA (pre-miRNA) occur in the nucleus and is catalyzed by Drosha enzyme with its indispensable cofactor called DGCR8 in vertebrate and Pasha in invertebrates. Drosha enzymes contain usually two RNase III domains and a C-terminal dsRBD. The cleavage product has a 5′-phosphate and a 3′-hydroxyl with a characteristic 2-nt 3′-overhang. The 2 nt 3′ overhangs of pre-miRNA hairpins are recognized by Exportin-5/Ran-GTP complex and then exported to the cytoplasm (210). Subsequent step of metazoan miRNA biogenesis occur in the cytoplasm with Dicer enzymes. pre-miR-155 is one example a product of Drocha cleavage **fig. 1.2**.

Dicers In plants miRNA biogenesis fully occur in the nucleus and is catalyzed by Dicer enzymes (like the second cytoplasmic step in metazoans). Dicer enzymes weight about 220 kDa and contain from N-terminus to the C-terminus, at least a PAZ domain and two RNase III domains. In mammals and plants Dicer(s) are more complex than in other multicellulars and contain an helicase domain, a PAZ domain, a domain of unknown function called DUF283 that is very likely to be a non canonical dsRBD, two RNase III domains, and one or two C-terminal dsRBDs. There is only one C-terminal dsRBD in mammals and usually 2 in plants **fig. 2.1** . Apart from mammals, Dicer is often present in multiple variant and in that case usually each one is predominantly devoted to a pathway (e.g. siRNA *vs.* miRNA). Dicer PAZ domain region has a pocket that recognize the heterogeneous 3'overhang (183; 180) and the 5' phosphorylated end interacts with a conserved pocket around the PAZ domain (213), which seems to be the most important recognition pocket. The three dimensional distance from the 5' phosphorylated end to the RNaseIII domains determine the length of the Dicer product. The 5' pocket is conserved in Dicers with a pre-miRNA cleavage activity like in human and *Drosophila* Dicers and likely in other higher eukaryotes; it is not conserved in Giardia and fungi, unknown in plants (40; 213). There are usually multiple double stranded RNA binding partners (dsRBD) for each Dicer, containing two to three dsRBD. SiRNA processing from long dsRNA precursors usually require more than 2 Dicer cleavages, versus one for canonical microRNAs in metazoans and two in plants. Thus, contrary to microRNA, high number of siRNA can be generated from a siRNA precursor

2.1.3 Loading of the small RNA on the messenger RNA(s) and dsRBDs

RISC and RLC Eventually one or both strand of the dsRNA, called the miRNA or siRNA targets a messenger RNA (297; 141) with an interaction mediated by the complex responsible for small RNA-directed RNA interference, called the RNA-Induced Silencing Complex (RISC). The RISC complex is the effector complex of RNAi. The core of RISC is composed of a dsRBD free Ago protein and a small silencing RNA. In the case of miRNA, the word "miRISC", miRNA induced silencing complex is used and "siRISC" with siRNAs. The RISC becomes functional after the release of one of the strand of the miRNA/miRNA* duplex, usually miRNA*. Loading the mi/siRNA on the mRNA require an unwinding of the duplex and is performed by the RISC loading complex (RLC), which is usually a subset of the RISC and contain the Ago protein. The RLC apparently contains the Dicer/cofactor/dsRNA complex thus contains the same dsRBD as the ones in the Dicer complex, although their effects might be different. In addition to the protein involved in miRNA biogenesis, the RISC complexes contain additional proteins that won't be described here. Similarly, there are various additional dsRBD free proteins that regulate the miRNA biogenesis pathways. They include P53 in human, hnRNPA1 in *C. elegans* for instance. The known ones don't have dsRBD and thus are not included in this review.

MiRNA and siRNA extracellular signaling: systemic and more Once the mi/siRNA is isolated, RNAi can occur in the cell or elsewhere since it is systemic i.e. mi/siRNA can propagate from cell to cell. Thus mi/siRNA effects are not limited to the cell that produces it. The nature of the potential proteins members involved in this process is still unclear. RNAi in plants and nematodes can be involved in extracellular signaling pathways (274). It appears to be an efficient mechanism for these organisms that use RNA interference as an antiviral mechanism since from a localized contact with the virus, the defense is spread to other parts of the organism. Studies in mammals, using others antiviral mechanisms, are actively performed to investigate whether a similar mechanism could exist and their potential effects (development, antiviral...). By now, it has been found that in mammals, miRNA can be actively secreted from tissues and present in stable entities in the serum, reviewed shortly in (300). A siRNA transporter called SID-1, homologous to the *C. elegans* SID-1 protein involved in siRNA transport has been discovered in humans (61). The transport is bidirectional and has been shown in the case of microRNA-21 to be an important factor for the resistance to gemcitabine in human adenocarcinoma cell (68). Potential roles of dsRBDs in these process in not known. Surprisingly, miRNA regulation can occur between very distinct clades as a recent study found that food intake (300) can regulate mRNA expression of the "predator": rice, containing MIR168a can inhibit LDL receptor aptamer protein 1 expression in human liver.

2.1.4 Variety of dsRBD containing proteins involved in mi/siRNA biogenesis

Thus, RNA interference appears as the last step of a progressive biogenesis pathway always orchestrated by dsRBD containing proteins and by Ago proteins. Four types of proteins contain dsRBDs in miRNAs and siRNAs pathways **fig. 2.1** :

- in ADAR enzymes containing multiple dsRBDs (3 in each human ADAR1 and ADAR2)

- in RNaseIII enzymes (Drosha and Dicers) at their C-termini in one or two copies, and possibly one in the DUF283 region.

- in the cofactor of RNaseIII enzymes, with a common topology of two to three repeated dsRBDs. DsRBD containing cofactors (TRBP, PACT, HYL1, etc) are involved in multiple steps of the small silencing RNA biogenesis and their role is potentially not the same along the pathway. In the light of the diverse data on dsRBD containing cofactor, it can't be excluded at the time of writing that these proteins might be linked to an unknown/unrealized phenomenon.

- to this list can be added proteins that interact with the cofactors and also possess dsRBDs and mediate an important interplay with other pathways, which is the case of the interferon-induced double-stranded RNA-activated Protein Kinase R (PKR) in mammals.

2.2 Roles of dsRBPs in mi/siRNA biogenesis in some well characterized multicellulars organisms

This is the biological functional context of the work performed in this thesis. A simplified overview of the mi/siRNA pathways is given in **fig. 2.2**.

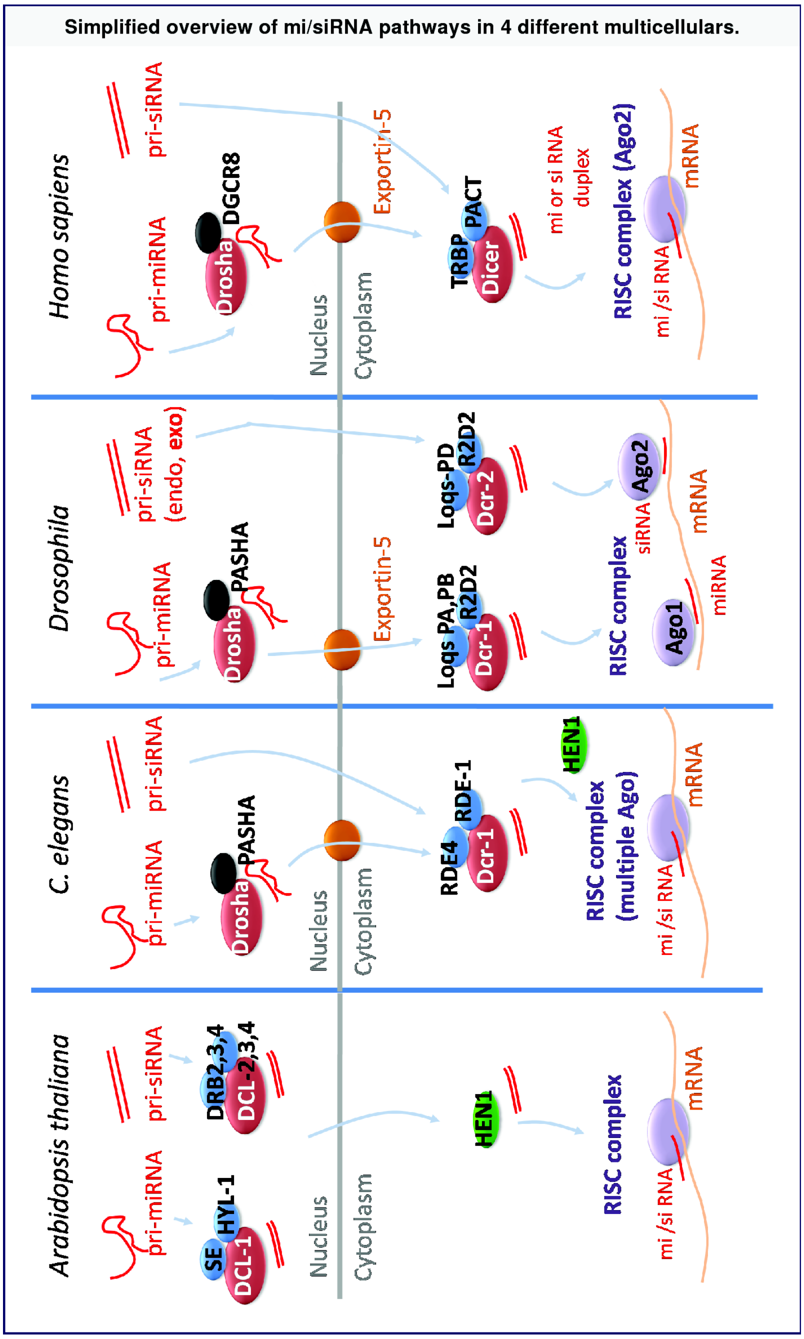


Figure 2.2: Simplified overview of the siRNA and siRNA biogenesis pathways in *Arabidopsis thaliana*, *C. elegans*, *Drosophila melanogaster* and *Homo sapiens*. There are many subtle features and uncertainties in these pathways and the reader is referred to the main text for more details about the protein names and functions.

2.2.1 dsRBPs in *Arabidopsis thaliana* mi/siRNA biogenesis

Types of small silencing RNA pathways in *Arabidopsis thaliana* MiRNAs, exo- and endo-siRNAs pathways exists in *Arabidopsis thaliana*. RNAi is used for development and regulation (miRNA), chromatin structure (endo-siRNA) and anti viral defense (exo-siRNA).

In plants, contrary to metazoans, small silencing RNA biogenesis occurs in the nucleus. The model system *Arabidopsis thaliana* possesses multiple Dicer-like enzymes and each small silencing RNA precursors is differentially targeted by a subset of DCL proteins, a feature that seems to have been evolutionary selected to face the multiplicity of plant viruses (19). MiRNA primary precursors are produced by the RNA polymerase II. The generated transcripts have a stem-loop structure usually longer than the one in mammals. Primary endo-siRNA depends on the RNA polymerase IV and V (289), associated with DRB2 and DRB4, two dsRBD containing cofactors with antagonistic effects (62). A dsRBD free RNA dependent polymerase is associated to siRNA production

Dicers-like in mi/siRNA biogenesis: two to three dsRBDs *Arabidopsis thaliana* has four Dicer Like (DL) ribonucleases, DCL1, DCL2, DCL3 and DCL4. Each of them has major roles but despite these trends, the DCL proteins are partially complementary. DCL1 participate in microRNA biogenesis (150). DCL2 produces virus-derived siRNAs as an antiviral mechanism and endo-siRNAs (95), and DCL3 produces endogenous siRNAs (289). DCL4 is the main processor of endogenous RDR6-dependent trans-acting siRNAs, a subgroup of piRNAs (95) and of virus derived siRNAs (52). Additionally a dsRBD in the former DUF283 region, *Arabidopsis thaliana* Dicer like ribonucleases have 2 dsRBDs at their C-termini called DCLn-A and DCLn-B respectively with n in $\{1,3,4\}$, apart from DCL2 with only one dsRBD.

DRB cofactors in miRNA biogenesis: pairs of dsRBD Four dsRBD containing double stranded RNA binding Proteins (DRBs) cofactor assist miRNA biogenesis in *Arabidopsis thaliana*: DRB1 (also called HYL1), DRB2, DRB3, DRB4 and DRB5. A pull down assays (and far-western assay using denaturing gels) study has proposed that each member of the DRB family interacts directly and specifically with one partner of the DCL family (e.g. HYL1/DCL-1 ; DRB-4/DCL-4) (122), a property that would suggest a specific modulation of the DCL family by their respective cofactors. As with Drosha in mammals, hairpin contained in the pri-miRNAs are processed by a complex composed of DCL1 (two C-ter dsRBD) and its cofactor HYL1 (two dsRBD) as well as the zinc finger protein Serrate (SE) (176; 181). Both HYL1 and SE are required for an accurate processing of DCL1 and they accelerate the cleavage rate of some pri- and pre- miRNAs (57). DRB2, DRB3 and DRB5 have their expression restricted in the shoot apical meristem region (62) and function in non canonical microRNAs pathways in *Arabidopsis thaliana*. DRB2 is important for miRNA biogenesis in the shoot apical meristem region (62) where the two cofactors HYL1 and DRB-2 are involved in a complex regulation of the biogenesis. DRB2 is either antagonist or synergetic to HYL1 (DRB1) depending on the microRNA (62). HYL1 is localized mainly in the nucleus and has a putative nuclear signalization at its N-terminus. DRB2 is localized in the cytoplasm (122). DRB3 and DRB5 are not involved in the miRNAs processing steps but are required to mediate RNAi of non canonical miRNAs associated with DRB2 (64).

DRBs have a pair of dsRBD at their N-terminal side **fig. 2.1**. Plant cofactors do not have a C-terminal dsRBD but a sequence without homology to other known proteins to date, with repeats of residues (28 res repeats for HYL1)

HEN-1 in mi/siRNA biogenesis: 2 dsRBDs in a methyltranferase In *Arabidopsis thaliana*, the methyltransferase Hua Enhancer 1 (HEN1), with 2 dsRBDs catalyze the methylation of the 2′-hydroxyl group at the 3′ end of the miRNAs and siRNAs before the RISC step.

Thus about 8 dsRBDs are involved in plant miRNA biogenesis and 4 in siRNA biogenesis. The involvement of HYL1 or a similar dsRBD containing protein in *Arabidopsis thaliana* siRNA biogenesis is unclear. Enlightening this ambiguity would be interesting to understand the specificities of miRNAs *vs* siRNA biogenesis in *Arabidopsis thaliana* and the role of HYL1 dsRBDs.

2.2.2 dsRBPs in *Caenorhabditis elegans* mi/siRNA biogenesis

Types of small silencing RNA pathways in *C. elegans* The *C. elegans* genome encode at least 112 microRNAs with no obvious phenotype in deletion mutants, thousands of endo-siRNAs that act on mRNA transcripts related to spermatogenesis and transposons, exo-siRNAs and a specific class of 21U RNA (235). The last class is characterized by a 21 nucleotide long RNA with a 5′ uridine-monophosphate and conserved in other nematodes (235). The deletion mutations in 83% of *C. elegans* miRNAs genes shows that the related miRNAs are not essential for its development and viability without major abnormal phenotypes (196). This has been interpreted as significant redundancy in miRNAs or the pathways they regulate.

Particularities of the miRNAs and siRNAs pathways in *C. elegans* *miRNA*. Canonical miRNA biogenesis in *C. elegans* is performed from RNA polymerase II primary transcripts 5′ capped and polyadenylated with a stem loop structure. The Microprocessor complex is composed of Drosha-1 and Pasha-1 (DRSH-1 ad PASH-1) (82). The stem loop pre-miRNAs of about 70 nt has bulges in its stem. After export to the cytoplasm, it is cleaved by a Dicer protein, DCR-1, also involved in siRNA pathway. Four mirtron miRNAs are also directly processed by DCR-1 in the cytoplasm (82). *siRNA*. DCR-1 has one dsRBD containing cofactor called RDE-4 (260; 214) and the biogenesis of exo-siRNAs require another protein, RDE-1 (82). *C. elegans* DCR-1 forms a complex with RDE-4 during the first steps of RNAi, from long dsRNA to short siRNA duplex precursors. RDE-4 is however not required for the development of *C. elegans* (260). DCR-1 partner RDE-4 means RNAi defective-4, a property that from which derives its name; it is required for the production of siRNAs (260). The processing of endo and exo siRNAs is rather complicated is still unclear (82). There are multiple types of Argonaute proteins in *C. elegans*, not reviewed here (82).

Topology of the dsRBD containing protein involved in miRNAs and siRNA biogenesis in *C. elegans*. *C. elegans* DCR-1 (1910 residues) has a predicted typical metazoan domain organization, with a predicted dsRBD fold between the helicase domain and the PAZ domain. The siRNA related DCR-1 cofactor RDE-4 (385 residues) is a protein of about 44 kDa and is predicted to have a tandem of dsRBDs in the N-terminal part and a long C-terminal region (260).

2.2.3 dsRBPs in *Drosophila melanogaster* mi/siRNA biogenesis

Types of small silencing RNA pathways in *Drosophila melanogaster* RNAi in *Drosophila melanogaster* is mainly devoted to viral defence, reduction of the spread of transposons, structured loci and overlapping transcripts. RNA Polymerase II synthesise the primary transcript of canonical microRNAs. It contains a hairpin with imperfect base pairing, contrary to plants. Additionally, some miRNA precursors in *Drosophila melanogaster* are produced from mirtrons (211; 236). MiRNA precursors are exported in the cytoplasm by exportin 5. An endo-siRNAs pathway exist in Drosophila (43). Endogenous siRNAs are produced from hairpin transcripts or dsRNAs from mobile elements (92). RNAi is an efficient anti-viral mechanism in Drosophila (114). The precursors of exo-siRNAs derive from dsRNAs appearing during viral replication or by biotechnologically introduced long dsRNAs or hairpins. siRNA biogenesis generates siRNAs of about 21 nucleotides whereas miRNAs are comprised between 21 and 24 nucleotides (92). The topology of the dsRBD containing proteins involved in miRNAs and siRNA biogenesis in *Drosophila melanogaster* is given in **fig. 2.1** and the main biological roles of these proteins are given in the next paragraphs.

Drosha/Pasha Canonical miRNAs are cleaved in the nucleus by Drosha with its indispensable cofactor Pasha (54) producing a 60-70 nt stem-loop product with 2 nt-3′ overhang extremities. The canonical pri-miRNA has a 7-methyl guanosine cap and poly(A). siRNA precursors are not dependent of Drosha/Pasha.

Dcrs Two Dicer-like proteins, Dicer-1 (Dcr-1) and Dicer-2 (Dcr-2), are involved in miRNAs and (endo, exo)-siRNA biogenesis respectively (268; 168). Dcr-1 does not have a functional DExD/H domain so miRNA biogenesis is not ATP dependent. Contrary to Dcr-1, Dcr-2 has a functional DExD/H domain, and ATP hydrolysis is required for the processing of long siRNAs, and seems to have the ability to use its helicase domain to cleave internal regions of long siRNAs (277). So both Dcr-1 and Dcr-2 alone have the ability to process pre-miRNAs, the absence of the DExD/H domain in Dcr-1 restricts it to the processing of pre-miRNA. The Dcr-1 and Dcr-2 cleavage product of pre-miRNAs and pre-siRNAs respectively are sorted into two distinct Ago enzymes, Ago-1 and Ago-2 respectively.

dsRBD containing cofactors: R2D2 and Loqs There are four dsRBD containing cofactor in Drosophila melanogaster: R2D2 (172), and three Loquacious (Loqs) isoforms: Loqs-PA, Loqs-PB, Loqs-PD. R2D2 and Loquacious isoforms are paralogs (37% similarity and 25% identicity in the dsRBD regions (92). R2D2 and Loqs act at the same level than human TRBP and PACT. The different roles of these dsRBD containing cofactor are particularly complicated and not yet fully clear. The dsRBD cofactor R2D2 is required in both Dcr-1 and Dcr-2 pathways (212). Loquacious proteins are generated from a primary transcript by alternative splicing: Loqs-PA (3dsRBDs), Loqs-PB (3dsRBDs) and Loqs-PD (2 dsRBDs). The third dsRBD of Loqs present in Loqs-PA and Loqs-PB is a non canonical dsRBD and is required for the interaction with Dcr-1 (296; 84). Loqs-PD is the isoform specialized in the endo-siRNA pathway (306; 119). Loqs-PD has a special 22 residues carboxy-terminal sequence that is required for the interaction with Dcr-2 (120; 197) and not Dcr-1 (306); maybe the site, a part of it, allows the rest of the protein to fold as proposed in (197).

MiRNAs production is associated to Dcr-1. Although Dcr-2 alone has the ability to cleave both pre-miRNAs and pre-siRNAs as previously mentioned, inorganic phosphate and R2D2 inhibit pre-miRNAs cleavage (30). Loqs-PA and PB are associated to Dcr-1 for pre-miRNAs processing. They both contain three dsRBDs as their homologue TRBP and PACT. Loqs-PD is dispensable for miRNA biogenesis. Loqs-PB increase miRNA biogenesis, by an unknown mechanism (better affinity, enzyme turnover, conformational change, etc). The presence of Loqs-PB alters the size of miRNAs, generating miRNA isoforms sufficiently different to have distinct targets that the ones generated without Loqs-PB or generated with Loqs-PA.

SiRNAs production is associated to Dcr-2 and the cofactors R2D2 and Loqs-PD. R2D2 and Loqs-PD act sequentially in the siRNA pathway. First, Loqs-PD is involved in the processing of the double stranded RNA with Dcr-2 (increase affinity, although Dcr-2 might work without it) (187). Dcr-2 interacts with R2D2 and Loqs-PD, forming the RISC-loading complex (RCL). R2D2 prevents Dcr-2 from processing pre-miRNAs (30). Then R2D2 is required for Dcr-2 to load endo-siRNAs or exo-siRNAs into Ago2-RISC (83; 266). Loqs appear dispensable for the miRISC assembly (174). R2D2 prevent the loading of siRNA precursors on Ago-1 (83). However other studies have concluded that R2D2 is dispensable for siRNA production (173), including endo-siRNAs so these molecules appear as potentially non dispensable guides of the small RNA biogenesis.

RISC Mismatches in the central region of miRNA/miRNA* duplex with an initial uridine nucleotide in the miRNA favor the integration on Ago-1 whereas good base pairing in the central region of the siRNA duplex with an initial cystidine favor the integration in Ago-2 (92). Ago-2 RISC silence mRNA with high complementarity. The loading of miRNAs and siRNAs on RISC is helped by Hsp90 and Hsc70. For siRNA, after degradation of the passenger strand, the methyltransferase Hen-1 adds a 2′-O-methyl group to the 3′ end of the guide strand (125).

2.2.4 dsRBPs in *Homo sapiens* mi/siRNA biogenesis

Types of small silencing RNA pathways in *Drosophila melanogaster* As in other mammals, in *Homo sapiens* both miRNAs and endo-siRNAs pathways exist and contrary to many multicellulars, the RNAi is not directly used as an antiviral response. The antiviral response for dsRNAs longer than 30 nucleotides is mediated by the interferon pathway in which the dsRNA binding protein PKR is involved, which is however linked to the mi/siRNA pathway. Shorter dsRNAs (21 to 25 nt) induce the RNAi mechanism, a property that is used by some viruses encoding miRNAs. Other pathways include the mirtrons (279) and snoRNA derived miRNA (292; 198). The diversity of key proteins involved in mi/siRNA is much smaller than that of other metazoans like *Drosophila melanogaster* or plants ones.

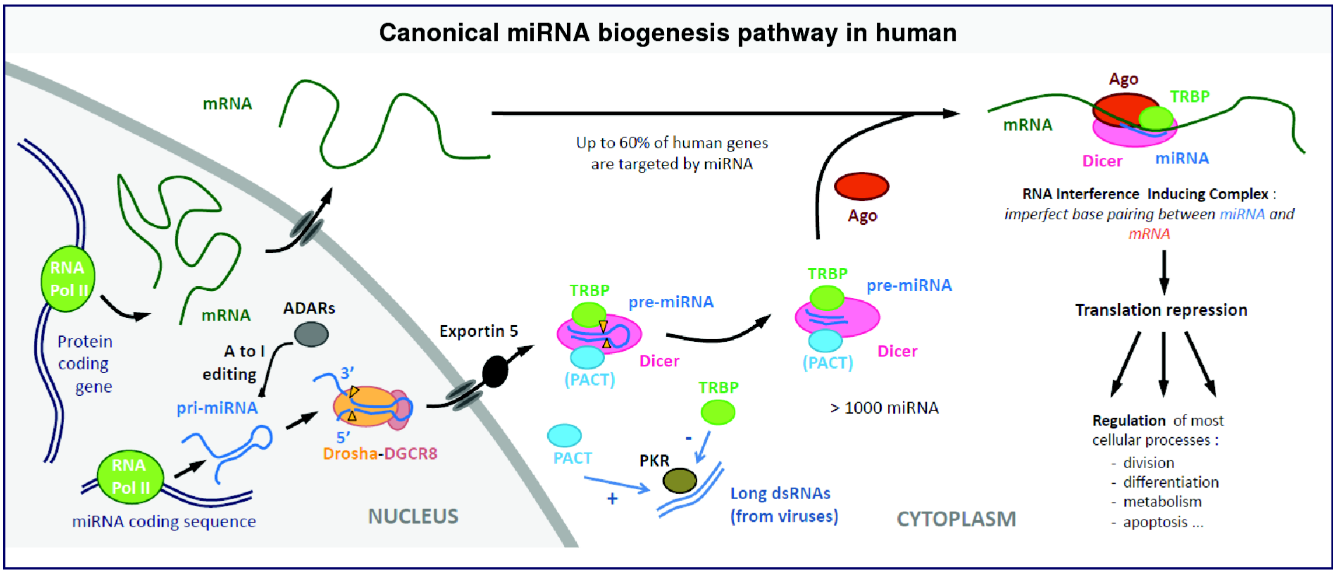


Figure 2.3: Illustration of the canonical miRNA biogenesis pathway in humans.

Microprocessor complex and canonical miRNAs From non coding RNA or mRNAs introns (142), polymerase II or sometimes III synthesise long 1-2 kbp primary transcripts (pri-miRNAs) that contain one or more 60-70 base pair (bp) stem-loop structure which includes the 20-25 nucleotides (nt) miRNA. Multiple hairpins can lead to distinct microRNAs. Intronic miRNAs could be generated before the splicing of introns (142). In the nucleus, most canonical pre-miRNAs are produced by the microprocessor complex. The enzymatic component of the microprocessor is the type 2 RNase III endoribonuclease Drosha (166) that interacts with a dsRBD-containing protein, DGCR8 (DiGeorge syndrome critical region gene 8) (111). DGCR8 binds to 10 bp of stem of the hairpin and is necessary for Drosha catalysis (111). Drosha enzymes may dimerize to form a heterotetrameric complex with DGCR8 (112). Similarly multiple copies of DGCR8 seem to be involved in the interaction with primary miRNA precursors (72). The product of the microprocessor is an imperfect hairpin of about 55 to 70 nucleotides (nt) in length containing/characterized by a 3′ two nucleotide overhang **fig. 1.2**. The 2 nt 3′ overhangs of pre-miRNAs hairpins are recognized by Exportin-5/Ran-GTP complex and then exported to the cytoplasm (210). A representation of the biogenesis of canonical miRNAs in humans is given in **fig. 2.3**.

Mirtrons and other non canonical miRNAs Recently high-throughtput sequencing in mammals has demonstrated the existence of microprocessor independent processes that can generate canonical miRNAs like precursors (9): mirtrons and functional RNA derived miRNAs. Mirtrons are non canonical pre-miRNAs generated by splicing of short introns with a hairpin potential (279). The splicing product is a lariat which has the 30 branch-point is ligated to the 50 end (279). After cleavage of this product by a lariat debranching enzyme, the mirtron can fold as a hairpin. This product have in majority 1:1 overhangs (17) can then be exported to the cytoplasm by exportin 5, as canonical pre-microRNA (211). Recently, a subclass of mirtron that does not require splicing by Drosha has been discovered (121).

Human Dicer has at least 3 know cofactors: manganese ions are required for the catalysis and two proteins TRBP and PACT have been shown to be members of the Dicer complex (presented in the next paragraph). The subcellular localization of the Dicer complex is mainly the cytoplasm but there are recent evidences of presence of Dicer and TRBP in the nucleus, whose function is unknown (7). Interestingly, Dicer is found in association with ribosomal DNA chromatin in mammalian cells (249). Electron microscopy analysis has revealed a characteristic L-shape (155). Dicer has a helicase/ATPase region but it is ATP independent (226; 298). The efficiency of Dicer processing seems to depend strongly on the structure of the substrate RNA. Human Dicer cleaves faster pre-miRNAs that dsRNAs *in vitro*, suggesting that the recognition and/or the processing of these substrates might be different. It has been shown that maximal cleavage rates with a Michaelis Menten model differed by more than 100 fold (178).

Eight domains have been identified on human Dicer, which are from N- to C-terminal: a region called DExD/H helicase comprising an helicase ATP binding domain with a DECH box and a Helicase C-terminal region, a central dsRBD (55) (formerly called DUF283 region as in other Dicer enzymes), a PAZ domain, two RNaseIII domains and a C-terminal dsRBD. Mutations that prevent cleavage of one of the RNase III domains have no effect on the cleavage capability of the other (299) showing that the two catalytic sites in the RIIIa/RIIIb function independently of each other. The Helicase region of Dicer is dispensable for RNA processing.

Dicer cofactors: TRBP and PACT Dicer is believed to interact with two homologous proteins, HIV-1 transactivation response (TAR) element binding protein TRBP (35; 109; 148) and simultaneously or alternatively with PACT (167; 148). Both TRBP and PACT are reliably predicted from sequence homology to be composed of 3 dsRBDs, which has been proved for the first two dsRBDs of TRBP (290) and the first dsRBD of PACT (RIKEN, PDBid 1dix).

TRBP. TRBP is homologous to PACT (42% identity) and to *Drosophila melanogaster* Loqs1 (233). There are two isoforms of TRBP resulting of splice variants, called TRBP1, TRBP2 (109). TRBP1 and TRBP2 differ at their N-terminus by 22 more residues in TRBP1. The functional difference between these two isoforms has not been reported. High sequence conservation exists between the C-terminal dsRBD of TRBP, PACT and loqs (109). TRBP/Dicer interacts independently of the presence of dsRNAs (109). The exact roles of TRBP and PACT still remain not perfectly clear. They are both dispensable for the activity of Dicer *in vitro*. TRBP is important for both miRNAs and siRNA biogenesis, originally demonstrated in (35; 109) and reviewed in (233) which underlie some inconsistencies in both studies regarding the effect of TRBP depletion and its Dicer stabilization activity. Multiple studies performed from these original articles show clearly multiple effects of TRBP that will be detailed through this review. An electron microscopy surface of the human-Dicer bound to TRBP has been solved a 20 Åresolution by negative staining (156), showing the expected L shape of the enzyme but not allowing to assess the position of TRBP in the density. The Dicer-TRBP complex can be sensitive to Adenosine to Inosine editions. For pri-miR-155, such edition at two defined positions in the stem part close to the cleavage site replace an A:U by an wobble pair U.I which block its cleavage by Dicer so the precursor accumulates (137). A study performed on the protein TRBP in HeLa cells (and insect cells) has shown that TRBP exists as a mixture of multiple phospho-isoforms, the most abundant form being however the non-phosphorylated protein (215). Four serine phosphorylation sites exist in TRBP all of which are located in the inter-domain regions (215). Phosphorylation of TRBP by the mitogen-activated protein kinase (MAPK) Erk has been shown to stabilize the interaction between TRBP and Dicer but not significantly affect the dicing activity of the complex (31).

PACT. PACT shares 42% identity with TRBP. TRBP and PACT can interact independently of Dicer and dsRNA (148; 44). PACT is able to bind TRBP and Dicer, without requiring an intermediate RNA. Like TRBP, PACT is dispensable for Dicer cleavage of miRNA precursors (167). However, contrary to TRBP, PACT does not facilitate Dicer cleavage (167). Actually, PACT presence in the Dicing step of Dicer is unclear since a study (7) failed to detect PACT in immunoprecipitation of Dicer but found TRBP and Ago2. However, PACT is positively involved in RNAi mediated by siRNA (167). PACT associates to the RISC complex with Dicer, TRBP and Ago2 (167). Depletion of PACT reduces the efficiency of RNAi and affects accumulation of mature miRNAs (167). It not yet a clear whether there is a differentiated role between TRBP and PACT for the processing of the different classes of small silencing RNAs. Neither is clear whether TRBP and PACT are partially redundant in mi/siRNA biogenesis.

PKR and its relations with TRBP and PACT. The interferon-induced double-stranded RNA-activated Protein Kinase R (PKR) is a dsRNA dependent serine/threonine protein kinase. It targets multiple proteins including eIF2α which causes a reduction of protein synthesis, reviewed in (263). TRBP and PACT have opposite effects on PKR: TRBP inhibits PKR (45; 108) whereas PACT activates PKR (108; 220; 217; 216). PKR is a key protein in the initiation of the interferon network suggesting links between miRNAs maturation and antiviral response via TRBP and PACT. Induced interferons are pro-apoptotic. PKR is composed of two dsRBDs in tandem and a C-terminal kinase domain. Additionally, since TRBP can be phosphorylated on serines, a phosphorylation of TRBP by PKR is a possible additional regulational link between small RNA biogenesis and PKR. A stress-induced phosphorylation at serine 287 of PACT weakens its interaction with TRBP and this increase PACT activation of PKR (248). Thus PACT and PKR together with their relationship with TRBP blur the separation between mi/siRNAs and the interferon pathway. This link is interesting in the light of the frequent (and potentially ancestral) use of RNAi against viruses.

miRNAs editing by ADAR1 and ADAR2 In human two ADAR enzymes, ADAR1 and ADAR2, are involved in mi/siRNA biogenesis (295). ADAR2 possess multiple isoforms and locates in the nucleus, whereas ADAR1 shuttles between the nucleus and the cytoplasm. The edition of miRNAs has been shown to occur from the primary precursor. Since ADARs are present in both the nucleus and the cytoplasm, miRNA precursors are potentially edited at multiple steps during their biogenesis.

RLC and RISC: multiple dsRBDs The RISC loading complex is composed of Ago2, Dicer and TRBP (106). Dicer is a key enzyme in RLC (238) and RISC complex for miRNAs (106) and siRNAs (238). Depletion of Dicer impedes RNAi with siRNA duplexes (56). The miRNAs loading complex (miRLC) composed of Ago2 and Dicer is formed prior to their binding with a pre-miRNA (186). Importantly, it is not clear whether the Dicer/TRBP complex that processes pre-miRNAs exists independently to the RLC complex, despite the fact that Ago2 is not required for this processing. Multiple experimental data suggest it is likely that the complex responsible for pre-miRNAs processing and miRNAs loading on Ago2 (RLC) are the same (106; 186; 167) with a coupled processing of miRNAs precursor and loading on Ago2. These studies report pre-miRNAs processed by RCL complexes containing Ago2 proteins. Once the small RNA is loaded on Ago2, the components of the RCL disassemble. Both TRBP and PACT are members of RISC (167). RISC has about 10 fold greater activity with the pre-miRNA Dicer substrate compared to siRNA duplexes (106). Recently, RNA helicase A (RHA) has been identified as human RISC-associated factor (232). Depletion of RHA reduces the concentration of active RISC and as a consequence diminishes RNAi, which underlines the importance of this enzyme as a RISC loading factor (232). RHA unwinds double-stranded RNA (and DNA) from 3′ to 5′. Human RNA helicase A has two dsRBDs in tandem on its N-terminal side (fig. 2.1 which lead to potentially 8 dsRBDs involved in human RLC/RISC. There are four Ago proteins in human (239) and they seem to be involved in miRNA biogenesis. Among them only Ago2 has an endonuclease activity for mRNA strands that display complementarity to the bound miRNA (called slicer activity). The association between some Ago proteins and a class of small RNA or a special silencing effect is unclear. Ago proteins do not have dsRBDs.

Chapter 3

Sequence and fold comparison of dsRBDs related to mi/siRNAs

This chapter focuses on the sequence and fold specificities of the dsRBDs in miRNA biogenesis. The general properties and variability of well characterized dsRBDs related to mi/siRNA biogenesis in multicellular organisms will be presented. As each new study of a dsRBD often shows new features as well as standard ones, the knowledge of the dsRBD group in multicellular is evolving quickly and various reviews have regularly aimed at underlying the common features and particularities of these domains (240; 188; 80; 264; 34; 58). The main general results will be synthesized in a first part. An original structural comparison analysis will then be performed on the dsRBD related to mi/siRNA biogenesis to systematically compare their dsRBD fold and to relate them to their function.

3.1 General features of double stranded RNA binding domains

3.1.1 Definition of double stranded RNA binding domains

The family of protein domains able to bind to double stranded RNA, now known as double stranded RNA Binding Domains (dsRBD), has been discovered in 1992 (254). The affinity of these domains to dsRNAs is actually not a sufficient or necessary property to define them. Other domains can bind dsRNAs like zinc finger domains or RNA recognition motifs for instance and some dsRBDs do not bind dsRNAs. Double stranded RNA binding domain are also called double stranded RNA binding motif (dsRBM). A few authors consider (mainly in PKR papers) that a dsRBD is the repetition of two dsRBM (305), but this is a minority in the literature since one dsRBM is enough to bind dsRNAs, since dsRBMs are not always in tandems and since they can often behave as independent folded structures. In 1995, the first structure of a dsRBD was reported (237): the second dsRBD from *Xenopus Leavis* Xlrbpa (a homologue of PACT) interacting with dsRNA. Xlrbpa is a 33 kDa protein similar to human PACT and TRBP that contains three dsRBDs. It is considered the reference dsRBD. Multiple NMR spectroscopy and x-ray studies from then have investigated the structures and properties of a multitude of similar domains. These data enable to defined canonical dsRBDs as domains "with an $\alpha 1 - \beta 1 - \beta 2 - \beta 3 - \alpha 2$

topology with α-helices formed on the same face of an antiparallel three strand β-sheet and able to bind double stranded RNA. Because of the variability of dsRBD, the dsRBD class borders is rather fuzzy and non canonical dsRBD terminology (sometime called dsRBD type B) is used for domains with the same fold and no dsRNA binding affinity or a moderate difference in the fold compared to common dsRBDs, like additional helices at the extremities or extended secondary structures.

3.1.2 Common features of canonical dsRBDs

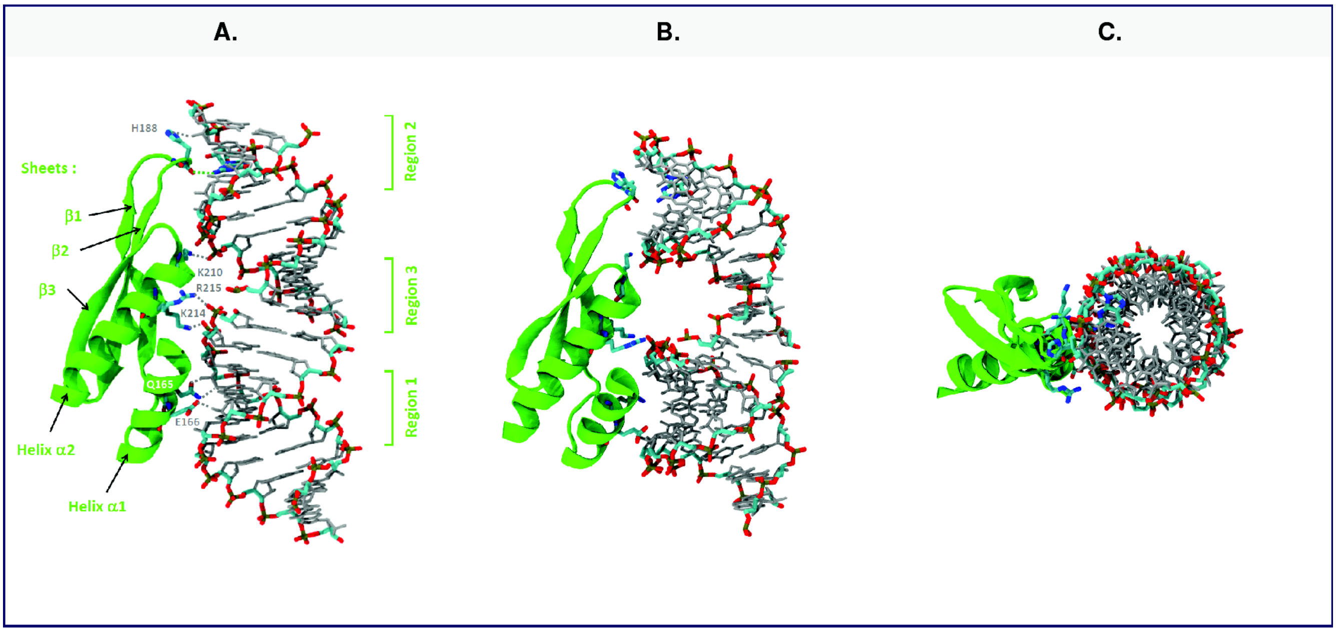


Figure 3.1: Example of the fold of a typical dsRBD, TRBP-dsRBD2 interacting with stacked 10 bp dsRNA (PDBid 3adl). The dsRBD is represented in green and the main residues involved in the interaction with dsRNA are represented as tubes colored according to their atomic composition (carbon: light green, oxygen: red, nitrogen: blue, phosphor: ochre). The dsRNA covalent bounds and atoms are represented as gray tubes apart from the regions interacting with the protein, colored: the phosphate backbone and a guanine, represented like the protein interacting residues. **A.** Side view of the dsRBD/protein complex. Hydrogen bounds between the dsRBD and the dsRNA are represented with gray dashed lines. The name of the main interacting residues, the names of the secondary structure elements and the position of the three dsRBD RNA interacting regions are given. **B.** 35° inclination of the A view from the top towards the reader. **C.** 90° inclination of the A view from the top towards the reader.

Structural data on canonical dsRBDs interacting with dsRNAs have invariably reported common features of the interaction that were reviewed in multiple articles (80; 264; 34; 58; 240): The crystal structure of the interaction between TRBP-dsRBD2 and two short 10 bp dsRNAs stacked collinearly is given as illustration in fig. 3.1.

- an elongated fold that binds dsRNAs so that the main axis of elongation of the two molecule are almost parallel. The domain do not wrap around the dsRNA helix. Only a small part of the dsRNA is involved in the interaction, potentially enabling multiple dsRBDs to bind close to each other.

- the preferred nucleic acid binding partner of canonical dsRBDs is dsRNA but some dsRBDs have lower affinities for ssRNA, dsRNA and RNA/DNA hybrids.

- both the dsRBD and perfect A-helix dsRNA usually undergo very limited structural changes upon binding.

- the typical dsRBD interaction surface on dsRNA corresponds to about 16 base pair i.e. 1.5 turn of RNA helix. However, dsRBDs seem to be able to accommodate smaller interaction sites. Outside of the small silencing RNA proteins, it has been shown that apical loops of hairpins can interacts with some dsRBDs (285; 275). A group solved by NMR spectroscopy a structure of *Saccharomyces cerevisiae* Rntp1 with a snoRNAR47 stem-tetraloop of 14 base pair (285), which is slightly shorter than the canonical dsRNA recognition site by dsRBD. They concluded that the tetraloop is recognized and the importance of this interaction has been demonstrated with *in vivo* experiments. However the moderate size of the physiological RNA does not prove that the loop site is preferred compared to an RNA with long dsRNA regions. The structure shows that the first helix has an orientation very favorable for the interaction with the minor groove of the loop of its target RNA. Absence of interaction with an apical loop has also been observed in the structure of ADAR2-dsRBD2 where the $\alpha1$ helix is not favorably oriented for an interaction with the minor groove of the loop, but present with ADAR2-dsRBD1. There is a lack of information on the behavior of most dsRBDs involved in mi/siRNA biogenesis regarding potentials hairpin apical loop interactions.

- the loop lengths are variable, which makes the dsRBD chain length variable, between 65 to 90 residues. Some dsRBDs have an additional helix in C-terminal but so far such dsRBDs were not reported in proteins related to mi/siRNA biogenesis.

- dsRBDs interact mainly with non sequence dependent parts, with the phosphate backbone and 2′OH groups of the ribose mediated by hydrogen bounds or by water. Interactions with base rings have been detected in some structures (188), which will be presented in **section 4.2.2**. Three binding regions have commonly been reported on dsRBDs interacting with dsRNA, called region 1, 2 and 3 following the N-terminal to C-terminal sequence. Region 1 corresponds to a side of the first short helix α-1. It is neutral or slightly negatively charged, usually with a conserved glutamate. It interacts with the minor groove of the dsRNA helix via the 2′OH of the ribose sugars. Region 2 comprise the loop between stand β-1 and β-2, is neutral or slightly negatively charged with a small positively charged part, commonly a GPxH motif with an important histidine residue. This region binds to dsRNA minor grove, especially with the 2′OH groups of the ribose sugars. Region 3 is the N terminus extremity of helix 2, positively charged with a conserved KKxAK motif (x: any residue), and binds to major grove RNA via direct or water mediated contact with the phosphodiester backbone. Along the dsRNA helix, the region appears in the order (1,3,2) or (2,3,1). Despite many dsRBD/dsRNA structures were solved with stacked 10 bp dsRNAs, which could introduce some biases, structure with longer dsRNAs show the same interaction interface. A detailed review of the many residues involved in the dsRBD/dsRNA interaction is presented in (188).

- Flexible loops. The flexibility of loop 2 has been detected for each dsRBD on which relaxation NMR studies have been performed (204; 229). This was also detected in the bound state for the third dsRBD of *Drosophila staufen* and a hairpin RNA (229). Additionally, in the crystal state, the flexibility of the second loop of Xrlbpa dsRBD2 has been detected. The loop 2 is very flexible for DGCR8-dsRBD1 and the dsRBD of Drosha (284). Loop 1 is often flexible but the flexibility seems not required for dsRNA binding: it is not found flexible in the 10^{-8}- 10^{-12} s^{-1} time scale by NMR spectroscopy relaxation in DGCR8-dsRBD1 whereas it is flexible in the unique dsRBD of Drosha that does not bind *in vitro* with a primary construct of miRNA 16-1 (284).

- the sequence is mainly conserved in the last third of the dsRBDs chain (i.e. C-terminal side). This corresponds to the largest α helix which interacts via hydrophobic residues with one side of the β-sheet and with the helix $\alpha 1$. These interactions form the hydrophobic core of the dsRBDs and appear fundamental to maintain the dsRBD fold. For dsRBDs that binds dsRNA, residues in the RNA binding region 1, 2 and 3 are also conserved. So called non canonical dsRBDs have usually good sequence conservation in the helix $\alpha 2$ and diverge more elsewhere.

- environment of the dsRBD. dsRBDs are always associated with other domains which often include other dsRBDs. In the data available up to now, the dsRBDs form mainly an independent fold in proteins and have usually a flexible linker of variable length that attach them to other domains.

3.2 Fold comparison of dsRBDs involved in mi/siRNA biogenesis

3.2.1 Selection of a set of well characterized dsRBDs involved in mi/siRNA biogenesis

Only the dsRBDs for which structural information is available will be given consideration here. Previous reviews on dsRBDs have already underlined the sequences specificities with multiple alignments as well as the regions important for interaction with dsRNAs. However no studies have concentrated on a systematic structural dsRBDs fold comparison. In the past 10 years, the number of dsRBDs structures solved in miRNAs and siRNA biogenesis pathways has highly increased, enabling to propose an original comparison of their fold. Because of the definition of dsRBDs using a fold rather than a well defined sequence motif, the search of the set of known dsRBD structures in miRNAs and siRNA biogenesis was performed with a three dimensional topology search. For the purpose of this study, the second dsRBD of TRBP with advantageous canonical structures solved in the apo and RNA holo form will be taken as reference. The solution structure of the second dsRBD of TRBP in the apo form (PDBid: 2CPN, (290)) was used as a template and the dsRBDs structures were obtained using the DALI server (124). Then dsRBDs of proteins involved in miRNAs or siRNA biogenesis pathways, with Z-score over 2 were selected and sorted whether the dsRBD is in the apo or holo form regarding its interaction with dsRNAs.

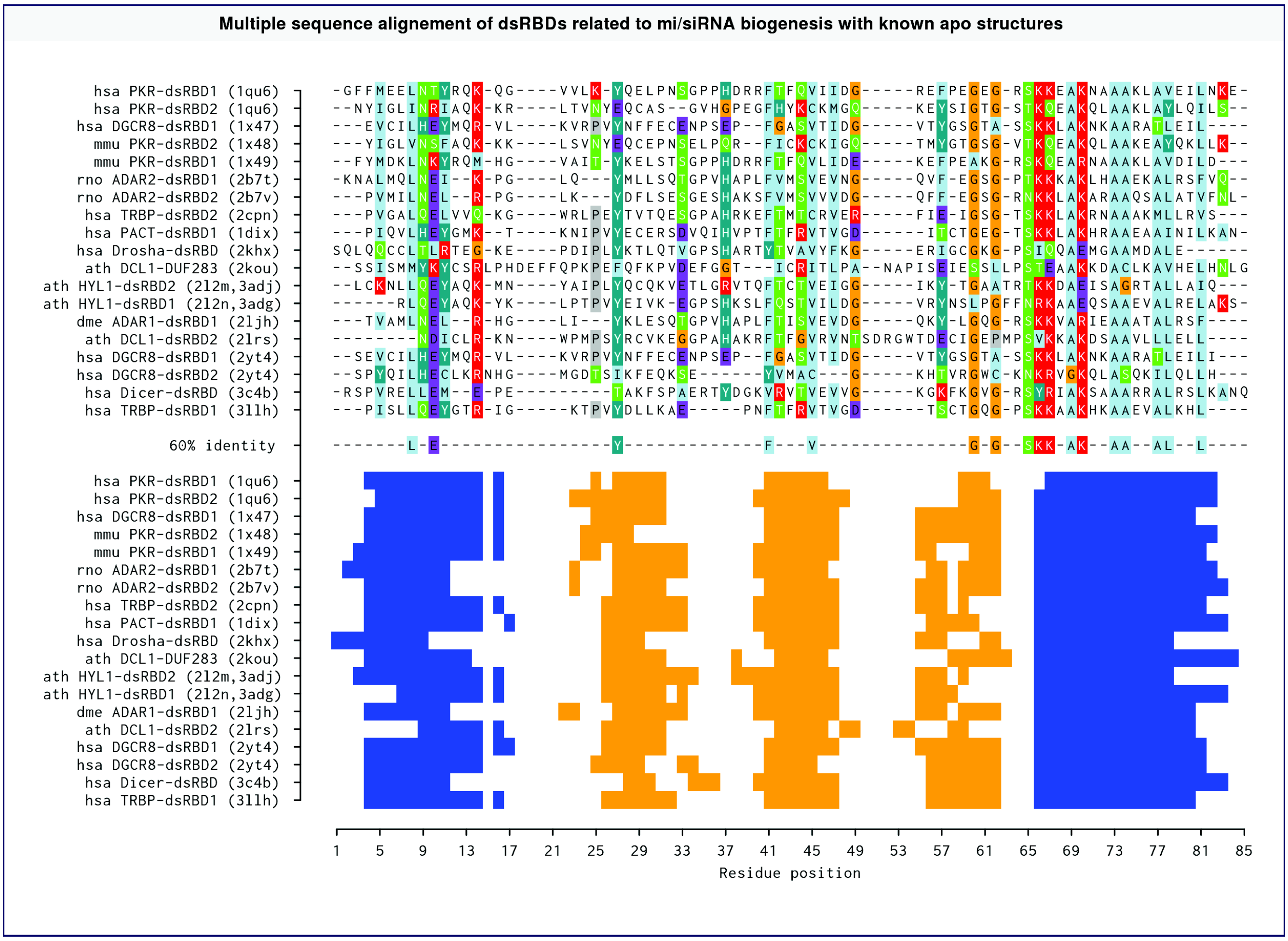
As already mentioned, despite not directly involved in the small silencing RNAs processing, the protein PKR that can interacts with TRBP and PACT and with 4 solved dsRBDs structures was included in the set. 20 PDB structures corresponds to dsRBDs in the apo form, 2 of which are almost identical and 2 contain 2 dsRBDs so 21 dsRBDs structures are available in relation to the miRNAs/siRNAs pathway. Among these 21 dsRBDs structures, 3 corresponds to the same dsRBD and were obtained with different methods or within a different protein construct. 8 PDB files correspond to holo forms of dsRBDs mainly with short dsRNAs stacked collinearly. More dsRBDs folds were studied by in solution by NMR spectroscopy (19 dsRBDs) than by crystal by x-ray diffraction (11 dsRBDs). Despite that for some dsRBD like TRBP-dsRBD2, the structures of both apo and holo forms are very similar, this might not be a general rule. Thus, since more data are available for dsRBDs in apo form, the comparison was performed with this set of data. To facilitate the comparison, the parts of the structures not included in the pure dsRBD region were removed from the PDB file. Using the Stride program to define the secondary structure (89), the criterion used was to select the subset of the PDB file between one residue before the start of the first helix of the dsRBD and one residue after the end on the second helix of the dsRBD. A summary of the different dsRBDs related to miRNAs and siRNA biogenesis available in the PDB to date is given in **tab. 3.1**. Solution structures of the first two dsRBD of RDE-4 have been solved but not yet added to the PDB (36). The sequence and fold features of this set of dsRBD is presented in the next two paragraphs.

3.2.2 Global sequence comparison of dsRBDs involved in miRNAs and siRNA biogenesis

Multiple alignments of the dsRBDs in relation to miRNAs and siRNA biogenesis in multicellulars that adopt an $\alpha - \beta - \beta - \beta - \alpha$ fold determined experimentally in the RNA apo was performed as detailed in the figure legend of **fig. 3.2**. The columns of aligned residue with a normalized entropy over 0.35 are colored according to the code defined in the figure legend. The result is given in **fig. 3.2** and shows some typical sequence feature of dsRBDs mentioned previously. In particular, the important sequence conservation in the long $\alpha 2$ helix is remarkable. This alignment is the basis of the following structural comparison.

Table 3.1: PDB structures selected for a fold comparison of dsRBDs involved in miRNA and siRNA biogenesis. The information id slitted for all dsRBDs contained in the structures. (*: the sequence of *Mus musculus* C-terminal dsRBD of Dicer is the same in *Homo sapiens* ; *Xenopus leavis* RBPA is also called PRKRA-b, like human PACT (PKRA). Holo annotations: (1): 10 bp dsRNA stacked collinearly ; (2): 22 nt, anologue of miR173/miR173* ; 3: GluR-2 RNA stem-loop.)

Organism	**Protein -dsRBD**	**PDBid**	**Apo *vs* holo**	**Context**	**Method**	**Reference**
Homo sapiens	TRBP-dsRBD1	3llh	apo	isolated	x-ray	(290)
Homo sapiens	TRBP-dsRBD2	2cpn	apo	isolated	NMR	(290)
Homo sapiens	PACT-dsRBD1	2dix	apo	isolated	NMR	RIKEN
Arabidopsis thaliana	HYL1-dsRBD1	2l2n	apo	isolated	NMR	(231)
Arabidopsis thaliana	HYL1-dsRBD1	3adg	apo	isolated	x-ray	(294)
Arabidopsis thaliana	HYL1-dsRBD2	2l2m	apo	isolated	NMR	(231)
Arabidopsis thaliana	HYL1-dsRBD2	3adj	apo	isolated	x-ray	(294)
*Mus musculus**	Dicer-dsRBD	3c4b 3c4t	apo	RNaseIIIb	x-ray	(60)
Homo sapiens	DGCR8-dsRBD1	1x47	apo	isolated	NMR	(201)
Homo sapiens	DGCR8-dsRBD1	2yt4	apo	DGCR8 core	x-ray	(252)
Homo sapiens	DGCR8-dsRBD2	2yt4	apo	DGCR8 core	x-ray	(252)
Homo sapiens	Drosha-dsRBD	2khx	apo	isolated	NMR	(201)
Drosophila melanogaster	ADAR-dsRBD1	2ljh	apo	isolated	NMR	(14)
Rattus norvegicus	ADAR2-dsRBD1	2b7t	apo	isolated	NMR	(257)
Rattus norvegicus	ADAR2-dsRBD2	2b7v	apo	isolated	NMR	(257)
Homo sapiens	PKR-dsRBD1	1qu6	apo	dsRBD pair	NMR	(203)
Homo sapiens	PKR-dsRBD2	1qu6	apo	dsRBD pair	NMR	(203)
Mus musculus	PKR-dsRBD1	1x49	apo	isolated	NMR	RIKEN
Mus musculus	PKR-dsRBD2	1x48	apo	isolated	NMR	RIKEN
Arabidopsis thaliana	DCL1-dsRBD1	2lrs	apo	isolated	NMR	(23)
Arabidopsis thaliana	DCL4-DUF283	2kou	apo	isolated	NMR	(227)
Homo sapiens	TRBP-dsRBD2	3adl	holo$^{(1)}$	isolated	x-ray	(294)
Xenopus leavis	RBPA-dsRBD2	1di2	holo$^{(1)}$	isolated	x-ray	(237)
Arabidopsis thaliana	HYL1-dsRBD1	3adi	holo$^{(1)}$	isolated	x-ray	(294)
Arabidopsis thaliana	HEN1-dsRBD1	3htx	holo$^{(2)}$	full protein	x-ray	(127)
Arabidopsis thaliana	HEN1-dsRBD2	3htx	holo$^{(2)}$	full protein	x-ray	(127)
Rattus norvegicus	ADAR2-dsRBD1	2l3j	holo$^{(3)}$	dsRBD pair	NMR	(256)
Rattus norvegicus	ADAR2-dsRBD2	2l3j	holo$^{(3)}$	dsRBD pair	NMR	(256)
Rattus norvegicus	ADAR2-dsRBD1	2l3c	holo$^{(3)}$	isolated	NMR	(256)
Rattus norvegicus	ADAR2-dsRBD2	2l3k	holo$^{(3)}$	isolated	NMR	(256)



```
hsa PKR-dsRBD1 (1qu6)          -GFFMEELNTYRQK-QG----VVLK-YQELPNSGPPHDRRFTFQVIIDG-----REFPEGEG-RSKKEAKNAAKLAVEILNKE-
hsa PKR-dsRBD2 (1qu6)          --NYIGLINRIAQK-KR----LTVNYEQCAS--GVHGPEGFHYKCKMGQ-----KEYSIGTG-STKQEAKQLAAKLAYLQILS--
hsa DGCR8-dsRBD1 (1x47)        ---EVCILHEYMQR-VL----KVRPVYNFFECENPSEP--FGASVTIDG-----VTYGSGTA-SSKKLAKNKAARATLEIL----
mmu PKR-dsRBD2 (1x48)          ---YIGLVNSFAQK-KK----LSVNYEQCEPNSELPQR--FICKCKIGQ-----TMYGTGSG-VTKQEAKQLAAKEAYQKLLK--
mmu PKR-dsRBD1 (1x49)          --FYMDKLNKYRQM-HG----VAIT-YKELSTSGPPHDRRFTFQVLIDE-----KEFPEAKG-RSKQEARNAAKLAVDILD---
rno ADAR2-dsRBD1 (2b7t)        -KNALMQLNEI--K-PG----LQ---YMLLSQTGPVHAPLFVMSVEVNG-----QVF-EGSG-PTKKKAKLHAAEKALRSFVQ--
rno ADAR2-dsRBD2 (2b7v)        ---PVMILNEL--R-PG----LK---YDFLSESGESHAKSFVMSVVVDG-----QFF-EGSG-RNKKLAKARAAQSALATVFNL-
hsa TRBP-dsRBD2 (2cpn)         ---PVGALQELVVQ-KG----WRLPEYTVTQESGPAHRKEFTMTCRVER-----FIE-IGSG-TSKKLAKRNAAAKMLLRVS---
hsa PACT-dsRBD1 (1dix)         ---PIQVLHEYGMK--T----KNIPVYECERSDVQIHVPTFTFRVTVGD------ITCTGEG-TSKKLAKHRAAEAAINILKAN-
hsa Drosha-dsRBD (2khx)        SQLQQCCLTLRTEG-KE----PDIPLYKTLQTVGPSHARTYTVAVYFKG-----ERIGCGKG-PSIQQAEMGAAMDALE------
ath DCL1-DUF283 (2kou)         --SSISMMYKYCSRLPHDEFFQPKPEFQFKPVDEFGGT---ICRITLPA--NAPISEIESSLLPSTEAAKKDACLKAVHELHNLG
ath HYL1-dsRBD2 (2l2m,3adj)    --LCKNLLQEYAQK-MN----YAIPLYQCQKVETLGRVTQFTCTVEIGG-----IKY-TGAATRTKKDAEISAGRTALLAIQ---
ath HYL1-dsRBD1 (2l2n,3adg)    ------RLQEYAQK-YK----LPTPVYEIVK-EGPSHKSLFQSTVILDG-----VRYNSLPGFFNRKAAEQSAAEVALRELAKS-
dme ADAR1-dsRBD1 (2ljh)        ---TVAMLNEL--R-HG----LI---YKLESQTGPVHAPLFTISVEVDG-----QKY-LGQG-RSKKVARIEAAATALRSF----
ath DCL1-dsRBD2 (2lrs)         --------NDICLR-KN----WPMPSYRCVKEGGPAHAKRFTFGVRVNTSDRGWTDECIGEPMPSVKKAKDSAAVLLELL----
hsa DGCR8-dsRBD1 (2yt4)        --SEVCILHEYMQR-VL----KVRPVYNFFECENPSEP--FGASVTIDG-----VTYGSGTA-SSKKLAKNKAARATLEILI---
hsa DGCR8-dsRBD2 (2yt4)        --SPYQILHECLKRNHG----MGDTSIKFEQKSE------YVMAC---G-----KHTVRGWC-KNKRVGKQLASQKILQLLH---
hsa Dicer-dsRBD (3c4b)         -RSPVRELLEM--E-PE---------TAKFSPAERTYDGKVRVTVEVVG-----KGKFKGVG-RSYRIAKSAAARRALRSLKANQ
hsa TRBP-dsRBD1 (3llh)         ---PISLLQEYGTR-IG-----KTPVYDLLKAE-----PNFTFRVTVGD------TSCTGQG-PSKKAAKHKAAEVALKHL----

60% identity                   -------L-E----------------Y------------F---V--------------G-G--SKK-AK--AA--AL--L----
```

Figure 3.2: Multiple alignment of the dsRBDs in relation to miRNA and siRNA biogenesis in multicellulars that adopt an $\alpha - \beta - \beta - \beta - \alpha$ fold determined experimentally in the RNA apo form up to 31 march 2013. The sequence alignment was performed with the MUSCLE program (65) which uses only sequence and not structural information. The Shannon's information theoretic entropy (246) was used with a 10 letter alphabet to measure the residue diversity with a classification close to the convention of Mirny and Shakhnovich (195) (Hydrophobic and Aliphatic [V,I,L,M], Aromatic [F,W,Y], Polar [N,Q], Positive [H,K,R], Negative [D,E], Tiny [A,G], Serine and Threonine [S,T], Proline [P], Cysteine [C] and Gaps [-]) and normalized using an implementation in the GNU R bio3D package (103). The color scheme is the following one: (A, V, M, I, L, F, C and W) are colored in light blue, (S, T, N and Q) are colored in green, D,E and R,K are colored in purple and red respectively, H and Y in blue-green and G are colored in orange. Aligned residues with a normalized entropy of conservation value below 0.35 are not colored. The type of secondary motif per residue follows the sequence alignment and was determined from the available structures using the Stride algorithm (89). Alpha helices are represented in blue and sheets in orange. *Species legend*: hsa: *homo sapiens*, ath: *Arabodopsis thaliana*, dme: *Drosophila melanogaster*, rno: *Rattus norvegicus*. As two structures are available for the first two dsRBDs of *Arabodopsis thaliana* HYL1, both are mentioned but only one per dsRBD was used (solution structures 2l2m and 2l2n). The bottom of the figure represents the secondary structure determined from the PDB files by the Stride algorithm (89).

3.2.3 Global fold comparison of dsRBDs involved in miRNAs and siRNA biogenesis

The comparison of the sequence of dsRBDs gives limited information on their diversity. The goal of this part is to compare their N-ter-α-β1-β2-β3 α-Cter fold. This is often performed using an all atom superposition of the structures with a minimization of the root mean square deviation (RMSD) between the structures. However since the RMSD weight each atom equally, deviation in any directions are considered the same so it is often hard to analyze accurately such superpositions when multiple substructures are variably positioned. Another possibility to compare the structures is to identify the most invariant region of the protein ("rigid core") which can be used to fit all the structures on it. This approach proved to be useful to identify some flexible regions that are likely to be related to the regions that are variable between different related proteins of distinct organisms. The superposition on a core approach was used here using the R package bio3D (103) implementing a previously described algorithm (98). The principle is to start form a sequence alignment of the dsRBDs and then align the structures using the sequence aligned residues as guides **fig. 3.2** . The ensemble of all residues fully aligned in the sequence alignment is considered as the core. The 3D structure ensemble is then averaged and the deviation of each atom to the average structure is estimated (98). The alignment position with the highest deviation is removed from the calculation and the structural averaging - deviation estimation - deletion is repeated until no residues remains. And eventually the part the proteins with deviation under a given volume are selected to define a core. A 2A^2 limit was used here. The superimposed structures are represented using VMD (128). The corresponding superposition on all the apo dsRBDs of **tab. 3.1** is given in **fig. 3.3**.

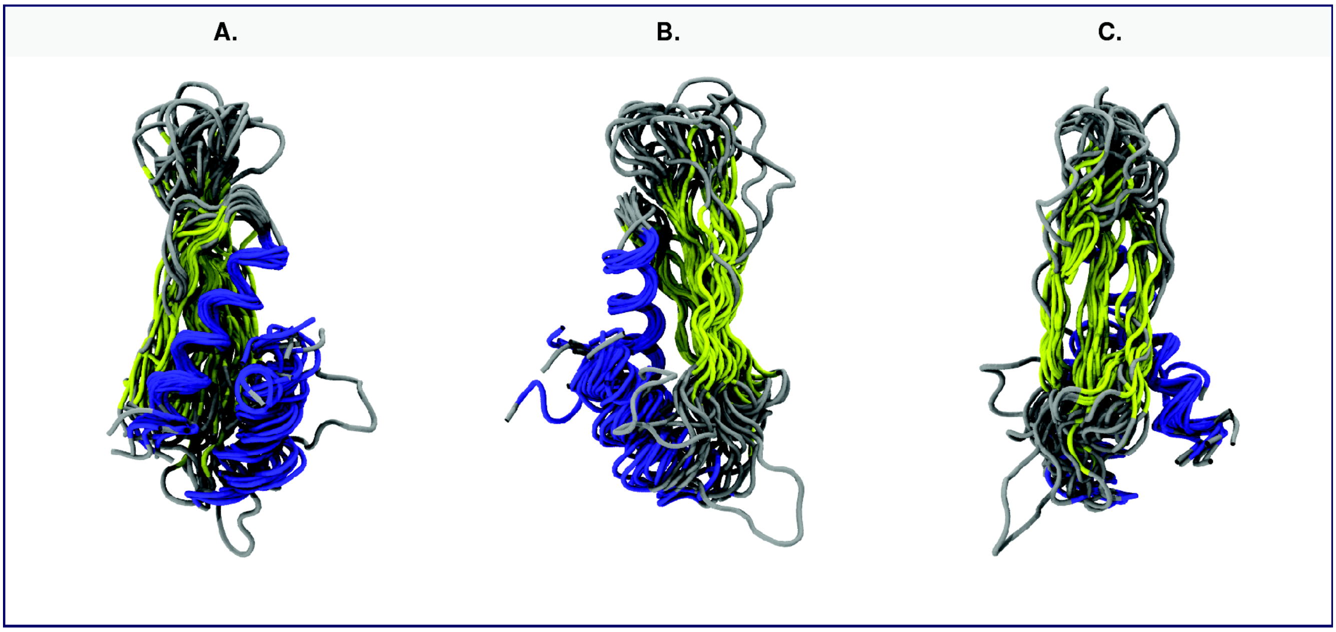


Figure 3.3: Result of the three dimensional alignment of the apo dsRBDs available structures from proteins related to mi/siRNA biogenesis in multicellulars listed in **tab. 3.1** The alignment strategy is explained in the text ; the structures are aligned on the long helix α2. Only the backbone of the dsRBDs are represented using tubes colored in blue for α helical secondary structure, in yellow for β sheets and in gray for loops. A, B and C represent different views of the alignment rotated around the dimension of higher dispersion of the dsRBDs atoms

Because of the variability of the dsRBDs structures, the structural part most common ("core") is reduced to 15 residues of the long helix 2 from positions 64 to 78 in the alignment and close to the small $\beta3$-$\alpha2$ loop. The comparison of the dsRBDs using this helix as a reference could appear artificial for such small protein domains but the analysis shows that this long helix is a key common structural feature among the dsRBDs and inspection of the structures show that the dsRBD fold is organized around this helix **fig. 3.3**. Both the β-sheet and the shorter N-terminal helix interact with the central C-terminal helix. The choice of this helix is also consistent with the high sequence conservation in this part of the protein.

The superposition shows a very high variability in the $\beta1$-$\beta2$ loop region and secondarily in the relative orientation of the N terminal helix compared to the C-terminal one. As many structures were solved by NMR, the diversity of the loop positions might just reflects lack of constrains in this region or irrelevant position constrained by the crystalization in the X-ray studies but it worth to note that the atomic position in these loops were not use in the classification process. The length of the N-terminal helix also appears to be variable (**fig. 3.2** & **fig. 3.3**). The $\alpha1$-$\beta1$ loop is also variable with 2 dsRBDs being extreme regarding this feature having unusual long loops (dsRBD of human Drosha and DUF283 domain of *Arabidopsis thaliana* DCL4). A simplified synthetic view of the variability of dsRBDs fold can be obtained by the realization of a partition of the structures based on the superposition. To distinguish as much as possible the folds, an exploratory statistic approach was used, with the workflow Principal Component Analysis (PCA) - Hierarchical clustering. The PCA is performed on a matrix which lines correspond to each dsRBD structure and which the columns corresponds to the repeated association of the x, y and z coordinates of the alpha carbon of residues fully aligned in the sequence alignment (i.e. no gaps in the sequence alignment, cf **fig. 3.2**). Importantly this removes the artifacts of the potentially unconstrained (possible in NMR) or overconstrained (possible in crystal) flexible loops as their sequence are very variable and correspond to regions not used in the alignment. The 37 fully aligned residues in **fig. 3.2** were used. Then using the PCA basis as a new set of discriminating variables, a hierarchical clustering can be efficiently performed. The results of this strategy are given in **fig. 3.4**. This approach should enable to identify the main structural features on which the structures differ in the context of the superposition performed. The higher the number of structure and their diversity, the more relevant are the results of this approach. We can notice that there is a lack of dsRBDs from RNaseIII enzymes involved in small RNA biogenesis which could lead to a change of the classes size and number. The 3 repeated dsRBD structure determined in different context enable to perform an internal validation of the method.

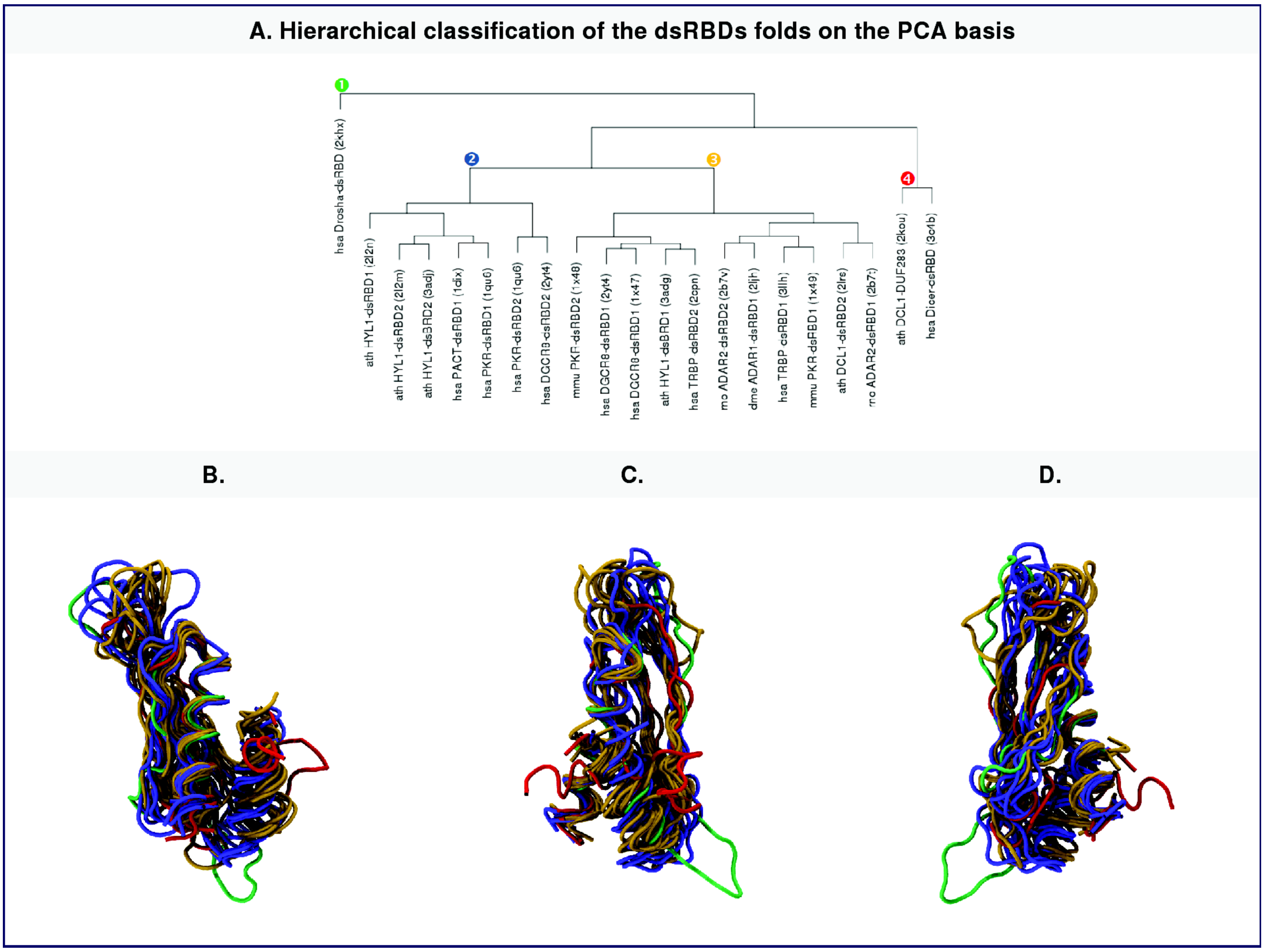


Figure 3.4: **A.** Hierarchical classification of the dsRBDs folds on the PCA basis. **B.**, **C.**, **D.**: Coloration of each dsRBD according to the group it belongs to, defined from the succession of: sequence alignment - structural alignment - PCA of residues fully aligned in the sequence alignment - hierarchical classification. (B.) The tree is thus colored to highlight the four groups. **B.**, **C.** and **D.** represent different orientation of the alignment along the axis of main dispersion of dsRBDs atoms using the same color scheme than **A.**

The hierarchical classification is well structured and suggested a relevant partition of the selected dsRBDs into 4 classes. The dsRBDs in the global superposition are colored according to their belonging to each class in fig. 3.4 to underline the differences.

Description of Group 1: human Drosha C-terminal dsRBD The C-terminal dsRBD of human Drosha (PDBid: 2khxD, (201)) is the first to be isolated with this analysis in a group because of its multiple unusual features. The structure of this dsRBD has been solved by NMR spectroscopy. As remarked by the authors of the structural analysis, this dsRBD has an unusual extended $\alpha 1$-$\beta 1$ loop negatively charged (see alignment), a unique dsRBD feature at the time of writing. This extension does not correspond to one the typical RNA binding region of dsRBDs interacting with dsRNA. The structural alignment shows also differences in the position of the $\beta 1$ strand. This dsRBD has the common sequence feature of dsRBDs and global charge distribution in the binding regions 1 and 2 but lack the common abundance of basic residues in region 3 (201). The functional effects of these features are not yet known. Human Drosha C-terminal dsRBD is required for pri-miRNAs processing by Drosha (111), however a study on the isolated dsRBD did not detected RNA binding properties (284). Interestingly, various studies have reported the lack of affinity for Drosha to its pri-miRNA substrate, which is linked to the requirement of its cofactor DGCR8 for canonical pri-miRNAs processing. This complex seems to possess unusual dsRBDs which could be linked to an unusual dsRNAs recognition, potentially linked to the abundance of bulges and mismatches in the canonical primary RNA. Further studies of the full complex would be required.

Description of Group 2 and 3: dsRBDs belonging to tandems dsRBDs of group 2 and 3 appear very similar and do not differ by large insertions in the loops. Some of the main differences are position and orientation of helix $\alpha1$, which appears variable. Additionally, the length of the $\beta2$-$\beta3$ are similar in this two groups but their position are different on average. The methodological influence of the position on these loops has been mentioned in the description of the classification method. It is however important to remark that these loops were not involved in the classification process. The two main differences identified between group 2 and 3 localize in two of the three RNA binding region of the dsRBD. And since the reference helix $\alpha2$ is taken as a reference for the comparison, this implies a difference of relative position of the 3 dsRNA binding regions. Most dsRBDs of available structures belong to these two groups. Importantly, the two structures of the first dsRBD of DGCR8 are found to be the closest and belong to the same group 3. Three of the four structures of the two dsRBDs of HYL1 solved by NMR spectroscopy and x-ray diffraction by two different groups localize in the second group. The second dsRBD of HYL1 by NMR spectroscopy and x-ray diffraction are found to be the most similar in this analysis. However the first dsRBD of HYL1 solved by x-ray is classified in the third group which indicates an effect related to the methodology so the belonging of this dsRBD to a group is not clear. Interestingly, apart for the mentioned case of HYL1 dsRBD1 and for DGCR8, all the dsRBDs belonging to a tandem have their partner that belongs to the same group. This is unexpected and could suggests the two main class of dsRBDs solved could correspond to two main classes of dsRBDs. It is however not straightforward to find obvious functional relations between the proteins belonging to a same group and as previously mentioned the structural differences are subtle. In particular, it is surprising to see that human and murine PKR dsRBDs belong to groups 2 and 3 respectively. On the contrary DGCR8 has its two dsRBDs in two different groups. This is interesting as the dsRBD pair of this protein is the most unusual among the ones reported. It is described in the part devoted to dsRBDs tandems (**section 4.3.1**). Group 3 possess a Dicer dsRBD fold that is sufficiently scarcely reported to be worth a comment. In *Arabodopsis thaliana*, DCL-1 can bind dsRNAs, like DCL-3 (122). In DCL1, the two dsRBDs were found to be essential for the microRNA processing. In the article reporting the solution structure of DCL1-B by NMR spectroscopy (23) the authors underlined the canonical dsRBD fold of this dsRBD with its particularities and interestingly, they found that DCL1-B is not specific to dsRNAs as it is able to bind dsDNAs. This is an uncommon property in the dsRBs group (23). It would be interesting to know what is the molecular basis for this property (fold/chemical features) and to analyze if these Dicer dsRBD could be involved in the growing reported association of Dicers with chromatin.

Description of group 4: two isolated Dicer dsRBDs Group 4 is very small and composed of two Dicer dsRBDs, *Arabidopsis thaliana* DCL1-DUF dsRBD and the C-terminal dsRBD of human Dicer. Another DCL-1 dsRBD, part of the C-terminal tandem is found in the group 3. The fold of these two dsRBDs is very unusual and characterized by a helix $\alpha1$ orientated between 45° and 90° to its position in other dsRBDs. A special orientation of this helix belonging to the dsRNA binding region 1 has already been mentioned outside of the small silencing RNA field in *Saccharomyces cerevisiae* Rntp1 as a way to facilitate interaction with loops (285). Interestingly, the role of the central DUF283 of Dicers, suspected to be a dsRBD is still elusive. The solution structure *Arabidopsis thaliana* DCL4 DUF has been solved by NMR spectroscopy (227) and indeed shows a dsRBD fold. In addition to obvious special orientation of the first helix, the authors also report a flexible 18 residue C-terminus, maybe a part of a linker with other DCL-4 domains and a lack of the conserved histidine in region 2. Interestingly, we can also note here that the two structure of this group have a reduced $\beta2$-$\beta3$ loop, corresponding to the dsRNA binding region 2. If the structures of these two isolated Dicer dsRBDs reflect the ones in the RNA holo form, it would strongly suggest a different way of interacting with the dsRNA substrate. DCL-4 DUF dsRBD has been reported to interact only very weakly with a perfect duplex dsRNA, the interaction being only detected by slight perturbation monitored by two dimensional NMR spectroscopy experiments (N-HSQC) with a 13 fold dsRNA excess and not by EMSA or by ITC (227). Interestingly, the deletion of the central dsRBD of human Dicer has only a limited impact of the pre-siRNA and pre-miRNA cleavage activity (178) which suggests it is not a key domain required for a global dsRNA affinity. Thus, the existence of this small group raises interesting question on the role of isolated Dicer dsRBDs.

The goal of this classification was mainly to present the different structures of apo dsRBDs and to find potential trends. Importantly, there is a lack of Dicer dsRBD in this classification. More structures would be necessary in the apo and holo form to see how stable the groups 2 and 3 are and to find a potential biological meaning or an origin of the fold differences. The fact the three dsRBDs in this comparison that does not belong to tandems have original folds (groups 1 and 4) might be coincidence but it appears attractive to investigate whether the "groups" 1 and 4 reflect undersampled classes of dsRBDs fold with specialized roles.

Chapter 4

Functionalities enabled by the dsRBDs in mi/siRNA biogenesis

Using the information presented in the three previous chapters, this chapter aims to compare the distinct roles of dsRBDs involved in mi/siRNA biogenesis with a strong link to the function of the protein that contains them and to the potential structural data available. dsRBDs related to mi/siRNA biogenesis are always associated with catalytic domains, either directly within the same polypeptidide chain (Dicer, Drosha, ADAR, RHA and PKR) or indirectly in cofactors interacting with these enzymes (TRBP, PACT, Loqs, DRBs etc). These dsRBDs add various functionalities to these enzymes (specificity, accuracy, localization etc). Despite many of these effects are likely to be interdependent, the present chapter tries to sort, potentially artificially, different roles of the dsRBDs in the mi/siRNA biogenesis pathways :

- dsRBDs in RNaseIII cofactors modulate enzyme activity
- mi/siRNAs precursor recognition properties added by dsRBDs and funneling roles
- effects of the arrangement of dsRBDs in tandems
- dsRBDs and protein/protein interactions

4.1 dsRBDs in RNaseIII cofactors modulate enzyme activity

Apart from a few cases like the requirement of DGCR8 for human Drosha, dsRBD containing cofactors are usually dispensable for the activity of their enzymes. Multiple experiments have shown however that these cofactors modulate the activity of the Dicer enzymes in different ways, including changes in the specificity and efficiency.

4.1.1 RNA dependent effects of dsRBD containing cofactors in Dicing efficiency in multicellulars

In various multicellulars, the dsRBD containing cofactors have a variable effect on the RNaseIII enzyme they are associated to. In *Arabidopsis thaliana*, HYL1 and Serrate (SE) stimulates of the amount of cleavage product for miR-167b with an enhancement between 2 and 20 fold (57). In *Drosophila melanogaster*, Loqs is not required for Dcr-1 to bind to miRNA/miRNA* duplexes

(174). As with other organisms, the effects on the detection of the cofactor Loqs is miRNAs dependent, either null or corresponding to a decrease of miRNAs production (174). In humans, the effect of TRBP on pre-miRNAs cleavage by Dicer is RNA dependent (31; 162). TRBP usually enhances Dicer activity of a few folds (2-10) but some negative effects on Dicer efficiency have also been reported for some precursors. The pre-miRNA dependent effects could be linked to a differential binding of TRBP, influencing the interactions of the TRBP/pre-miRNA complex with Dicer and/or a conformal change of Dicer (which is suggested to be autoinhibited by the helicase domain to which TRBP binds (178)). The structural basis of these differences remains elusive but is likely to be related to be special RNA recognition properties of dsRBDs, developed in this chapter.

4.1.2 dsRBP containing cofactor have an effect on the accuracy of Dicer enzymes in plants

In *Arabidopsis thaliana* where DCL1 products are highly heterogeneous regarding their length and sequence for a given precursors (which are long in plants), the addition of HYL1 and Serrate enable to have a well defined product corresponding to the physiological miRNA (57). DCL1 associates with HYL1 (113) for an accurate pre-miRNA processing (149; 57). HYL1 increased the accuracy defined with the expected product for miR-176b from 11% to 22%, and this reach 68% with serrate. These effects have been reported as a requirement for DCL-1 of these cofactors for an accurate processing. The related processes may be similar to the isomiRs ones, developed in **section 4.1.3** .

4.1.3 dsRBDs influence the isomiR production by Dicers in metazoans

The seed sequence of miRNAs determine the mRNA target. The seed region is located between nucleotides 2 and 8 in the mature miRNA **fig. 1.2**. For canonical miRNAs in metazoans, the role of Drosha and Dicer cleavage on the length of miRNAs depends first on which strand is the small silencing RNA that is to be loaded. If the miRNA is in the 5′-3′ strand, Drosha determines the 5′ end and Dicer determines the 3′ end, the opposite being true. Recently, various studies in different organisms have found that the presence of the dispensable dsRBD containing cofactors of RNaseIII enzymes could influence the length of the cleavage product, leading to different miRNA isoforms. These different miRNAs, differing by their 5′ or 3′ by one to two nucleotides in length, are called isomiRs. IsomiRs are differentially expressed in tissues and during development in *Drosophila melanogaster* (79) and in human (163), underlying that this process is very likely to be regulated. Importantly, the 1 to 2 nucleotide can be sufficient to change the seed sequence and/or the extremity of the small RNA duplex having the most stable base paring.

4.1.3.1 Loqs-PB and TRBP have an effect the isomiRs distribution

In *Drosophila melanogaster*, the dicing product of pre-miRNAs by Dicer-1 is affected by Loqs-PB. It is not the case for Loqs-PA, generating product of length identical to the Dicer-1 alone (91). The length of the miRNA with either Dicer-1 alone or Dicer-1 with Loqs-PA is mainly a 21 nt RNA. In the presence of Loqs-PB, Dicer-1 cleaves mainly a region two more nucleotide longer than without Loqs-PB. Since the miRNA-307a is in the 3′-5′ arm of the pre-miRNA precursor, Dicer-1/Loqs-PB processing leads to two more nucleotide at the 5′ end of the miRNA-307, leading to a change of the seed sequence. In the study, target genes have only been identified for the long isomiR form. Only the resulting iso-miRNA produced with Loqs-PB can target the predicted target mRNAs of glycerol kinase and taranis, a property verified *in vitro* for these two genes. As a result, Loqs-PB appears to be necessary for the production of the functional isomiR. A similar phenomenon has been found with miRNA-87, with the generation of a miRNA one nucleotide longer (91) and with miR-316, with one nucleotide shorter miRNA product. In the latter case, the miRNA corresponds to the 3′-5′ arm and the one nucleotide change does not result in a different seed sequence. This phenomenon is not systematic as the processing of pre-miR-let-7 *in vitro* leads to the same product independently of the presence of Loqs-PA or Loqs-PB.

In *homo sapiens*, this effect has been detected with the TRBP cofactor of Dicer but not with PACT (91; 162). The effect of TRBP is the generation of isomiRs with one nucleotide more than the canonical sequence. It has been reported for pre-miR-29, pre-miR-200a and pre-miR-34c and pre-miR-132-3p (162). For instance with pre-miR-132, both strands pre-miR-132-5p and pre-miR-132-3p can be produced *in vivo* and lead to distinct silencing. With Dicer alone there is a strong excess of pre-miR-132-3p, a 21 nt RNA whereas with TRBP, there is also a significant amount of 22 nt RNA product. In the latter case, both RNA have a distinct seed sequence. Interestingly, the change in sequence, changes the strand loaded on RISC, a phenomenon that is compatible in this example with a change of the extremity of the duplex with the most stable base pairing (162) (a phenomenon presented later in **section 4.2.1.2**). This example shows that even if the seed sequence is not modified, the change of 1 or 2 nucleotides at the extremities of the miRNA duplex can change the proportion of strand loaded on the RISC complex.

4.1.3.2 Towards a mechanistical understanding of Dicer cofactor effects on the isomiRs distribution

The first article reporting the cofactor effects on isomiRs production in *Drosophila melanogaster*, aimed at approaching the underlying mechanism. In the study, it was found using chimeras of pre-miR-let-7 and pre-miR-307a that the stem of the pre-miR-307a stem loop determines the possibility for Loqs-PB to affect the dicing position (91). As the Dicer product length is determined by the distance between the PAZ domains and the RNaseIII domains and since no real dsRBD sequence specificity have been found yet in this pathway, a model based on structural features on the precursor has been proposed in which Loqs-PB could either extend the pre-miRNA stem, leading to an elongated precursor stem and a shorter product (like pre-miR-316) either shrinking the stem, leading to a longer product (like pre-miR-307a and pre-miR-87) either with no changes in the stem length (like pre-miR-let-7) (91; 92). A key role of the internal loop is expected in this model; the loop could be either stabilized by the shrinking of the stem, or destabilized by the stem elongation. There is a need of high resolution structure of Dicer/partners/pre-miRNAs to clarify this.

The similarity between *Drosophila* Dicer-1 and human Dicer with one cofactor, Loqs-PA and PACT respectively, having no effect on the length of the product whereas the other one, Loqs-PB and TRBP respectively, has one seems interesting. Difference between these two pairs of cofactors could help to understand some requirements for the mechanisms. With two instances, one can only speculate on the reasons that underpin these properties. The main known common difference between [Loqs-PA, PACT] and [Loqs-PB, TRBP] is the length of the region between the two first dsRBDs and the last non canonical dsRBD. This region is longer in [Loqs-PB, TRBP] than in [Loqs-PA, PACT]. Additionally, TRBP has a longer interdomain region between the first two dsRBD. Assuming the prediction of the interdomain region as linker holds true, the longer linker of Loqs-PB could allow the dsRBDs to cover the surface of the precursor with fewer constraints than in [Loqs-PA, PACT]. It would be interesting also to know if there is a stoichiometry difference between [Loqs-PA,PACT] and [Loqs-PB,TRBP] interaction with some pre-miRNAs and affinity differences between the domains, and especially what is the extent of binding of the two first dsRBDs on the same precursor molecule: one dsRBD only binds, the first two binds, or even more if the stoichiometry is higher than 1. Certain dsRBD containing cofactor might be more efficient in stabilizing certain precursor structure than other cofactors.

These effects raise many questions: Is the ratio between Loqs-PA and Loqs-PB in *Drosophila* and PACT/TRBP in human a way to regulate in miRNA biogenesis? And more generally, what is the exact biological significance of these pairs of cofactors with distinct properties? In humans, TRBP and PACT are linked to other pathways, in particular the Interferon pathway where TRBP inhibits PKR whereas PACT activates it.

4.2 mi/siRNA precursor recognition properties added by dsRBDs and funneling roles

Apart from the canonical recognition of dsRNAs by dsRBDs presented in the Chapter 2, one can distinguish three additional feature of the RNA that can be recognized by some dsRBDs: the structure of different classes of small silencing RNA precursor, enabling to funnel some precursors in dedicated pathways, the hybridization stability in RNA duplexes and the ability to interact with specific bases in RNA duplexes.

4.2.1 dsRBDs and potential funneling roles of miRNAs and siRNAs classes

4.2.1.1 Various dsRDP cofactors contribute to miRNA *vs.* siRNA funneling

Some organisms have distinct sets of Dicer/cofactors, with clear trends in their substrate targets. dsRBP cofactors seem to have a key role in this sort of "funneling". In *Arabidopsis thaliana* DCL1 and HYL1 (i.e. DRB1) participates in the biogenesis of miRNAs and endo-siRNAs but among the multiple DCL/DRB complexes, it is the only complex reported to participate in miRNA biogenesis, suggesting that the complex has a special properties enabling the recognition of miRNA precursors. MiRNA biogenesis is partially impaired without HYL1, with a decreased formation of the pre-miRNAs (and misplaced cleavages) (149). This suggests that dsRBDs of HYL1 are a key factor in the recognition of miRNA precursors. Similarly, in *C. elegans*, RDE-4 is involved in the initiation of the RNAi pathway which may occur within a single complex, as suggested by the initial requirement of RDE-1 an Ago like protein for the association of Dicer-1 with dsRNAs (260). In the first biochemical study of RDE-4, no interactions have been found between RDE-4 and miRNA precursors (260). However RDE-4 binds dsRNA (260) in an apparently non sequence specific manner (214). This suggests that RDE-4 participate in the separation between the siRNAs pathway in *C. elegans* and the apparently less important miRNAs pathway. The underlying mechanism is not known but as suggested by a similar funneling process in *Drosophila melanogaster*, it may be related to a subtle interaction between the cofactor and its enzyme rather than a pure discrimination by the cofactor. Indeed R2D2 is essential to load endo-siRNA duplexes to Ago-2 and also to prevent their binding to Ago-1 (212). SiRNAs in *Drosophila melanogaster* are directed to Ago-1 in the absence of R2D2.

4.2.1.2 Human dicer C-terminal dsRBD as a potential contributor of miRNAs preference

The main difference between most pre-siRNAs and pre-miRNAs is the presence of a loop on pre-miRNA. Additionally in metazoans some mismatches and bulges in the stem of pre-miRNAs are not present in pre-siRNAs. Human Dicer can cleave both miRNAs and siRNA precursors. Kinetic studies have shown that human Dicer alone was able to process much faster pre-miRNAs than pre-siRNAs. In multiple turnover conditions, maximal cleavage rates of the Michael and Menten models of these two classes of precursors differs by more than 100 fold using classical double stranded siRNA precursors(31). Thus Dicer alone can discriminate between these two substrates. Human dicer C-terminal dsRBD is close the RNase III active sites suggesting a role in dsRNA binding (73). The deletion of this domain in full length human Dicer lead to a strong reduction of pre-miRNAs processing by human dicer (178; 299) and its deletion in a Dicer without the region including the helicase to the end of the PAZ domains also require the dsRBD for substrate binding (179). A study focused on the C-terminal dsRBD of human Dicer has shown that both long double stranded siRNA precursors and miRNAs hairpin precursors can interact with it with an affinity of the same magnitude (283). With perfect duplex resembling siRNA precursors substrate (33 bp) and product (22 bp), human dicer C-terminal dsRBD has a similar affinity which suggest that this dsRBD cannot distinguish between the substrate and product of Dicer and thus is not involved in the release of the product based on the length of the dsRNA (283). To mimic miRNAs, adding an octa-U loop to a duplex increased the affinity by a factor of almost 2 compared to the RNA duplex (283). It would be interesting to perform this comparison for more dsRNAs and miRNAs molecules. From these studies, the C-terminal dsRBD contribute to Dicer substrate binding and potentially slightly to a distinction between siRNAs and miRNA precursors. Potential mechanistic explanation of the difference observed awaits detailed structural data. The stoichiometry of the isolated C-terminal dsRBD of human Dicer with a dsRNA was estimated to be 1:1 with sedimentation velocity AUC (283). A crystal structure of a part of mouse Dicer containing the C-terminal dsRBD with 100% sequence identify with the equivalent domain in human showed a characteristic $\alpha\beta\beta\beta\alpha$ fold (60) with a few particularities. NMR relaxation studies have shown that the β1-β2 loop was not really flexible, which did not impaired high affinity dsRNA binding (with micro-molar affinity in the study conditions). Significant flexibility compared to the rest of the dsRBD was detected in β2-β3 loop and β3-α2 loop (283). β2-β3 loop is on the opposite side on the canonical dsRNA binding interface of dsRBD. The high flexibility in β3-α2 loop has also been detected in Drosophila staufen-dsRBD3 in the apo and holo state. Direct relationship between the loop dynamic and binding properties of the dsRBD remains to be determined. The miRNAs/siRNAs differential binding might occur in the context of the whole Dicer/cofactors/dsRNAs complex where cooperativity mechanisms with the different domains able to binds dsRNAs could exists. A similar comparison would be interesting with the cofactors of Dicer (TRBP and PACT). A study has already shown that TRBP could increase similarly the dicing rate of both pre-miRNAs and pre-siRNAs (31).

As mi/siRNA precursors are asymmetric molecules, many works have investigated the potential orienting effect of the dsRBD containing cofactor of RNaseIII enzymes. Indeed, the presence of a stem-loop structure and the apparent asymmetry in the base pairing stability of the two extremities of siRNA appeared as attractive orientation factors for the proteins.

In *Drosophila melanogaster*, R2D2 also binds to the more stable extremity of the siRNA duplex whereas Dcr-2 binds to the less stable extremity, allowing the RCL to asymmetrically load siRNA duplexes on Ago2 (267). Ago2 catalyses the degradation of the passenger strand and then cleaves the mRNA targeted thanks to its slicer activity (230). The two dsRBD tandems of R2D2 are both required for siRNA binding and for the asymmetrical loading of siRNA on RISC (173). The detection of an effect of the base paring stability difference on the strand loaded and in relation to a dsRBD containing cofactor has motivated the search of a similar phenomenon in other organisms.

HYL1 in *Nicotiana benthamiana* is involved with AGO1 for miRNA strand selection and directs guide strand selection based on the base paring stability asymmetry of the miRNA duplexes (63). It was observed that the strand with a 5′ lower thermostability was preferentially loaded. The mechanism is probably based on a preferential orientation of the miRNA duplex in the RISC complex triggering the miRNA* degradation (63). DCL4/DRB4 miRNAs and siRNA biogenesis seems to have a different mechanism (63), especially since the thermostability rules of the extremities does not seem to explain the strand loaded.

It has been suggested that TRBP can sense the asymmetry of siRNA in vitro (28). TRBP preferentially crosslinks to the 3’ends of the guide strands of siRNAs (105). However, Dicer and Ago alone seems have been reported to be able do this themselves. Human Dicer is able to sense the asymmetry and even to cycle if the orientation of the substrate is not ideal (207). The effect of TRBP which remains to be demonstrated for microRNA precursors could contribute the Dicer and Ago effects. Apart from the fact the more stable base paired extremity of small silencing RNA duplexes is expected to have a structure closer to a stable RNA A helix, structural determinants that underpin the preferential binding of some dsRBPs to these sites are not clear.

4.2.2 Some dsRBDs are dsRNAs sequence sensitive interactants

Among the dsRBDs involved in mi/siRNA biogenesis, various ones were reported to have base specific interactions. Commonly reduced to one for most dsRBDs, this feature appear very important in ADAR (ADAR2) dsRBD1 and dsRBD 2 according to an in solution structure of human ADAR2 with its natural substrate, complemented by mutational experiments (188). The sequence specific interactions in dsRBDs were reviewed recently in (188) and in (12) for ADARs. The case of the solution structure of the first dsRBD of human ADAR2 interacting with GluR-2 RNA stem-loop (256) is given as example in **fig. 4.1**.

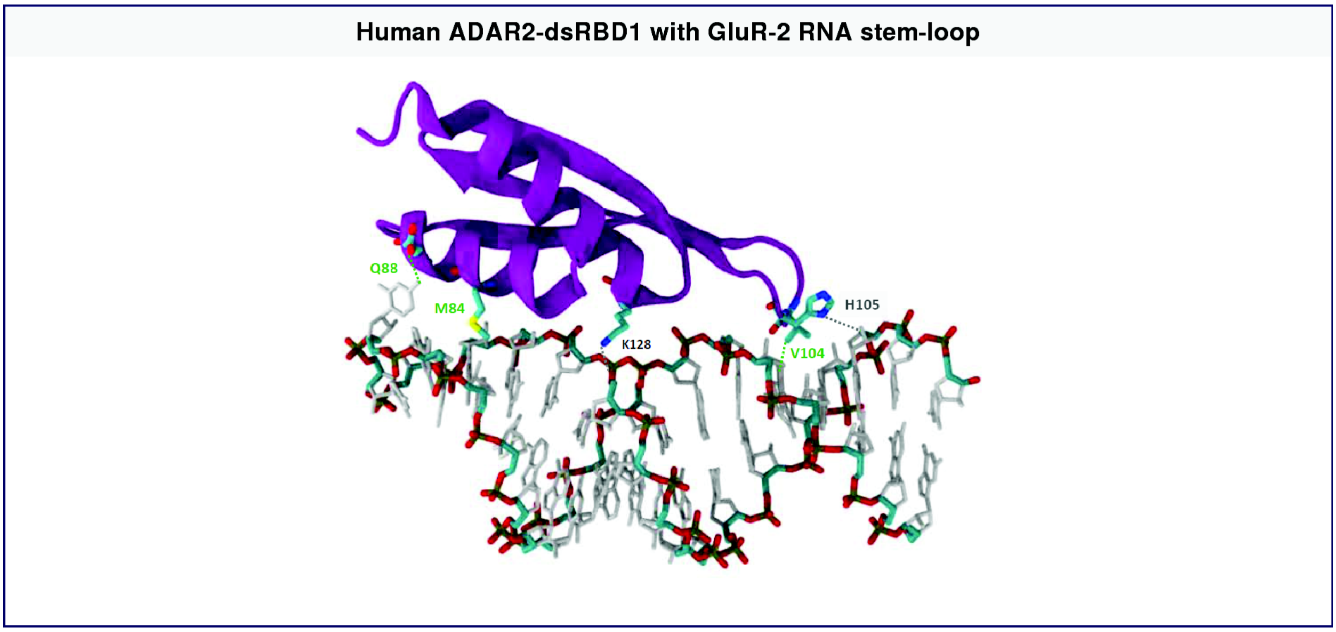

Human ADAR2-dsRBD1 with GluR-2 RNA stem-loop

Figure 4.1: Sequence specific interaction in the complex between the first dsRBD of human ADAR2 and the GluR-2 RNA stem-loop (PDBid: 2l3c ; (256)). The dsRBD is represented in purple and the main residues involved in the interaction with dsRNAs are represented as tubes colored according to their atomic composition (carbon: light green, oxygen: red, nitrogen: blue, phosphorus: ocre). The RNA is represented as a gray tube apart from the phosphate backbone. A hydrogen bounds between the dsRBD in the phosphate backbone is represented in gray (K128). Residues involved in specific interactions with the bases and their interactions are represented in green (Met84 and Val104: hydrophobic interactions; Glu 88 : hydrogen bound).

Sequence specific interaction with dsRBDs: only ADARs? Two regions of the dsRBD are involved in sequence specific interactions (188; 12) :

- helix $\alpha 1$. A methionine interacts with an adenine and a glutamine interacts with a guanine. These interactions that could not occur with different bases are thus sequence specific. This feature has been reported in the two dsRBDs of human ADAR2 (188).

- $\beta 1$-$\beta 2$ loop that interacts with the minor groove. This interaction is a hydrogen bound between the backbone carboxy group of a residue in the loop and the amino group of a guanine. This amino group is a specificity of guanine in the minor grove so the sequence specificity of this interaction. This interaction has been more frequently found than the one with helix $\alpha 1$ with different base pairs involving guanines (GC in two orientations compared to the axis of the dsRBD and GG): *Xenopus laevis* Xlrbpa, *Arabidopsis thaliana* HYL1-dsRBD1, human TRBP-dsRBD2, human ADAR2-dsRBD2 and human ADAR2-dsRBD2 with a GG pair (188). The importance of this interaction for the function of ADAR2 has been demonstrated by *in vitro* experiments by replacing the G-C base involved with A-U base pairs (256). These experiments showed a decrease of the affinity of the dsRBD for their RNA substrate by a factor of 1.5 and 6.5 fold for dsRBDs 1 and 2. *in vivo* experiments with double mutations in loop 2 residues involved in the interaction to change loop backbone structure in both dsRBDs decreased by about 80% ADAR2 editing (256). Another sequence specific interaction in the $\beta 1$-$\beta 2$ has been reported in *Arabidopsis thaliana* HYL1-dsRBD1 between the N3 atom of a guanine and a serine hydroxyl group (294; 256).

Multiple dsRBDs with sequence specific interaction capabilities like the two dsRBDs of human ADAR2 could increase the sequence specificity of the dsRBD containing protein.

Global sequence influence on dsRBDs binding: example of TRBP Taking into account the available data, important sequence specificity in all the proteins involved in the recognition and successive cleavage of miRNAs and siRNA precursors would be highly surprising. Indeed, there is a very large diversity of sequence of small silencing RNA precursors to be recognized and processed. However, a slight sequence effect could have been evolutionary selected to trigger some processing variation and on the other hand, some dispensable cofactor at one stage (TRBP, PACT) could also have a slight specificity not incompatible with the variable processing properties observed with them.

Interestingly, a moderate influence on the dsRNA sequence has been observed for TRBP. In the second published paper on TRBP (96), before any structural data, experiments performed on the peptide composed of residues 208 to 228 (position updated compared to the paper to take into account a plasmid LacZ fusion adding 39 additional residues) reported a much higher affinity for purely a G-C hairpin compared to its A-U counterpart. This region corresponds to the long α helix at the end of the second dsRBD (region 3), which is not involved in specific base interaction in the available structure of the full dsRBD. Whatever the actual interaction surface with this peptide which is the core of the second dsRBD, it is expected that the important KKxxK motif (x: any residue) has a strong contribution to the interaction and suggest a sequence influence of the sequence on its binding capacities. This could just correspond to a base pairing stability difference between the two hairpins and would require deeper investigations but is worth noting.

A large scale experimental study on 701 siRNAs has investigated the role of dsRNA sequence on TRBP binding by comparison of the binding affinities of multiple dsRNAs (135). They found that the nature of the residue at every $3n_{n \in \mathbb{N}}+1$ position in siRNAs correlate with TRBP affinity (and with RNAi activity *in vivo*). Adenines at positions $3n$ and $3n1+1$ and uridine at position $3n+1$ have a positive effect on RNAi. Guanine and cytosine bases at these positions led to a negative effect on RNAi. Analysis of the binding of 29 siRNA strongly suggested that TRBP has a key role in this property. It will be also interesting to know the potential role in this context of the sequence specific interaction between TRBP-dsRBD2 $\beta 1$-$\beta 2$ loop that interacts with a guanine in the minor groove. Precise positions of A and U might retrain the number of configurations of TRBP/siRNA complexes and lead to an increase of the affinity.

4.3 Effects of the arrangement of dsRBDs in tandems

When a succession of multiple dsRBD copies is present in a dsRBP protein, there is a systematic affinity difference for dsRNAs between the dsRBDs which usually corresponds to an association of a canonical dsRBD with a slightly less canonical dsRBD regarding the sequence. Importantly, the affinity relationship appears to be always the same for the different RNA analyzed in different works. The presence of multiple dsRBDs has been suggested to increase the affinity for RNA, the RNA specificity or to help the protein to bind simultaneously multiple dsRNAs. The latter possibility has however not been demonstrated so far. This part intend to present the most relevant data and to underline that the effects of this *a priori* simple organization are likely to be multiple and far from being clear.

4.3.1 dsRBDs in tandems are frequently independently folded units

All solution studies in dsRBD tandems, using liquid state NMR spectroscopy, have shown that the dsRBDs were independently folded, with no intramolecular interaction between these domains in the apo form. This has been shown for the first two dsRBDs of HYL1 (231), for human PKR (203) and human ADAR2 (256). There are thus no evidences for an obvious coherent increase of RNA binding surface. This flexibility could enable the dsRBD to bind multiple molecules or to bind with multiple dsRBDs on one dsRNA, which is the case for human PKR.

The structure of DGCR8 core in a crystal has been reported (252) and constitutes an exception among the literature on the fold of dsRBD in tandems. The structure includes the tandem of dsRBD with an arrangement resembling a two fold symmetry. The dsRBDs are compactly folded, interacts with each other and are linked by a small alpha helix with their putative RNA interacting surfaces oriented on opposite sides of the complex. If these properties also exist in solution, the two dsRBDs do not seem to be able to bind the same pri-miRNA or a simultaneous binding would require a bending of a pri-miRNA. Solution state experiments have aimed to explore this possibility, including a backbone assignment by NMR spectroscopy (234). The exact roles of DGCR8 dsRBDs which are fundamental for Drosha activity remains elusive but the structural difference between the two dsRBDs remarked in the fold classification suggests a different behavior than most of the dsRBD tandems.

4.3.2 Affinity differences between dsRBDs in tandems

dsRBDs in tandem with high affinity for dsRNA: mainly dsRNA affinty contributors? The dsRBDs with high dsRNAs affinity have systematically a canonical dsRBD fold and they are usually not reported to be involved in protein/protein interactions. The first dsRBD of HYL1 and the second dsRBD of TRBP are both dsRBDs with the highest RNA affinity in the tandem they belong to and have been extensively studied in solution and behave as monomers. Interestingly, domain swapping of the first dsRBD of HYL1 with the second dsRBD of TRBP has shown that the latter domain can supplant the one of HYL1 *in vivo* and *in vitro* (294). This suggests the main role of some high affinity dsRBDs in RNaseIII cofactor might be restricted to RNA binding.

In enzymes containing dsRBDs, the dsRBD with the weakest affinity for dsRNAs can even appear dispensable for the known function of the enzyme. The RNA editing enzyme ADAR1 of human and *Xenopus laevis*, has three dsRBD. In human, the third dsRBD is the most important in human for dsRNA binding in solution (175). It is also the case in Xenopus (22). The first dsRBD contribute to RNA binding, while the second dsRBD appear dispensable for ADAR function (175).

Structural investigation on the dsRNAs affinity differences between dsRBDs in tandems: example of HYL1 and TRBP Various structural and dynamical arguments have been proposed to explain the affinity differences between dsRBD in tandems. The case of *Arabidopsis thaliana* is interesting as extensive data from different groups are available. Two independent groups have investigated the structure of the two dsRBD of HYL1, at its N-terminus in crystal by X-ray diffraction (294) and in solution by NMR (231). The first dsRBD of HYL1 in a crystal has a canonical dsRBD fold (294) and in solution by NMR (231). A crystal structure of this domain with a 10 bp dsRNA has been solved (294). This domain has a fold very similar to the ones of the second dsRBD of TRBP and of *Xenopus leavis* dsRBD-2. Both HYL1 dsRBDs can interact with RNA but the affinity of the second dsRBD is much lower than the one of the first (122; 294; 231). Difference in affinity between the two dsRBDs of HYL1 are about a factor of 10 (1: 740 nM ; 2: 10 μM in (231)). HYL1 second dsRBD has a non canonical fold (294). In the crystal of apo dsRBD2, this dsRBD has significant changes in the regions 1 and 2 of the canonical dsRBD, compared to the second dsRBD of TRBP and of the first dsRBD of HYL1 crystallized with the same 10 bp dsRNA (294). The β1-β2 loop is conserved in the first dsRBD of the different dsRBPs in *Arabidopsis thaliana* (DRB1-DRB5) but not in the second one (231). This difference appears as a key factor in the RNA affinity difference between these two domains. The changes correspond to a different helix orientation in region 1 and a different loop configuration in region 2 (294). These remarks perfectly match with the global differences observed in the classification of the dsRBDs folds performed in **fig. 3.4** .

The structures of the first and second dsRBDs of TRBP solved by x-ray diffraction and NMR spectroscopy respectively have a very similar fold (group 3 of the fold classification). The second domain of TRBP has a cation-π stabilizing interaction which has been proposed to contribute to its higher stability and affinity compared to the first dsRBD, despite a sequence contribution is likely (290).

4.3.3 Tandems of dsRBDs as a way to cooperatively increase dsRBPs affinity for dsRNAs?

The dsRBDs in tandem with a weaker affinity for dsRNAs have been proposed in the discussion of many articles to be sensitive to structural feature of the RNA but this has not been demonstrated yet. In this case, some positive cooperativity would be expected for the binding of appropriate RNAs. More generally, various studies have aimed to characterize the cooperativity between multiple dsRBDs in tandem but the evidences reported are often insufficient to conclude clearly. One study report that *Arabidopsis thaliana* HYL1 dsRBD-1 has an affinity for dsRNAs about 40 fold weaker than the full HYL1 (294) whereas another one find that the affinity of a tandem construct with both dsRBDs is similar to the one of the first dsRBD (about 0.7 μM in this study) with a miRNA/miRNA* duplex (231). The difference between these two studies might reflect different behavior of the dsRBD on different RNA or reflect foreseen biases related to the different techniques used (ITC *vs.* EMSA respectively). A key difficulty in these analyzes is to adequately take into account the differences of stoichiometry between the experiments with dsRBDs in tandem and isolated dsRBD. In PKR the tandem of dsRBDs in PKR, the first dsRBD has the highest affinity for dsRNAs (265). The tandem dsRBD construct of PKR has an affinity 10 times higher than the affinity of the first dsRBD, which is also 10 times higher than the affinity of the second dsRBD (265). This relation suggests a positive cooperativity in the binding of PKR to the dsRNA analyzed. In TRBP, the two first dsRBDs bind to the same siRNA precursor molecule and seems to increase the global affinity for it, compared the single domains (290). This property could also have a role in the miRNA duplex length.

4.3.4 Tandems of dsRBDs as discriminators between mi/siRNA precursors and dsRNAs products?

A role of dsRBDs tandems in the discrimination between dsRNAs occurring along the pathway is conceivable. However, apart from the aforementioned possible case of the C-terminal dsRBD of human Dicer, no significant binding differences have been clearly reported along a small silencing RNA pathway. *Arabidopsis thaliana* HYL1 can bind pri- pre- and miRNA/miRNA* duplexes with similar affinities (294). In *C. elegans*, a study has proposed to compare the affinity of RDE-4 with long and small dsRNAs. In this study, RDE-4 has been reported to binds with more affinity to long dsRNAs, a decrease over than a thousand times in the (single) dissociation constant is reported in (214) comparing a 650 bp dsRNA with a 40 bp dsRNA, the latter being in the micromolar range. This has been interpreted as the ability of RDE-4 to distinguish between long and short siRNAs. Although mentioned by the authors themselves, the obvious difference of stoichiometry between the two RNA constructs (visible in the gel mobility shift analyzes) and its effect on the estimation dissociation constant have not been taken into account so the difference of affinity cannot yet be considered relevant.

4.3.5 dsRBD tandems as slicers on dsRNAs

The propagation of some dsRBDs on dsRNAs was reported recently, at the end of 2012, and at the time of writing, only one publication reports it using single molecule fluorescence analyzes (147). The studies were performed on three examples of dsRBD containing protein cofactors with dsRBDs organized in tandems: human TRBP and PACT and *Drosophila melanogaster* R3D1-L. These various examples enable to take this study into consideration here. Since the process is ATP independent, they called the propagation on dsRNAs a one dimensional "diffusion" process (147). Using single molecule fluorescence resonance energy transfer and one-color protein-induced fluorescence enhancement, the study reports that TRBP has a continuous and periodic movement from one end to the other end of Watson-Crick base paired dsRNA of length over 15 base pairs, moving along larger distances on longer RNA and probably moving along the entire length of the RNA. No diffusion was detected on DNA/RNA hybrid, DNA/DNA duplexes or on single stranded RNA. Interestingly, only the tandem of dsRBDs belonging to the N-terminal part of TRBP, known as the dsRNA binding region was found to be required for the diffusion process and each isolated dsRBD was not sufficient for the diffusion process to happen. The one dimensional "diffusion" process has also been detected for the Dicer/TRBP complex, the diffusing complexes having the ability to cleave the dsRNA (147) whereas human Dicer alone did not have the ability to diffuse on the dsRNA. The diffusion behavior was also observed with PACT and R3D1. Thus some dsRBD tandems can be grouped under a larger set of proteins that have the ability to diffuse on double stranded nucleic acids; the three examples reported in (147) are the first examples of diffusion on dsRNA. This one dimensional diffusion process could enhance the affinity of dsRBD for dsRNAs. It may also be one of the key roles of dsRBD containing proteins that are cofactor of RNaseIII enzymes in small silencing RNA biogenesis.

Despite the authors did not emphasize on differences between miRNA precursors and siRNA precursors, it is noteworthy that the average length of pre-miRNA precursor stem is 15-18 nt which is very close the minimal length for diffusion that they reported for dsRNAs. Additionally, the systematic motifs in miRNA precursors, bulges and loops, appear to impede the diffusion of TRBP. Thus the diffusion mechanism might be even more important for the processing of siRNAs related long dsRNAs or be a phenomenon avoided by the special propertied of miRNA precursors. The phenomenon could explain the increased or decreased efficiency of human Dicer associated with TRBP for some small RNA precursors. It will be very interesting to know why two dsRBDs are required and especially if the very common differences of affinity and dynamics observed in tandem dsRBDs is involved, as well as the potential role of the interdomain region. The relation between this diffusion property and other properties of TRBP remains to be elucided.

4.4 dsRBDs and protein/protein interactions

A common feature reported or strongly suggested in many dsRBD containing cofactors is the ability of the dsRBD possessing a non canonical dsRBD fold to interacts with other proteins, that belong to the small silencing RNA biogenesis pathway or not. In the miRNA biogenesis pathways, these interactions are mainly homodimerization, intramolecular interactions, interaction with other dsRBD containing cofactors and importantly with RNaseIII enzymes responsible for the small RNA precursor cleavage. Interactions with proteins not required for the biogenesis of small silencing RNA, are less characterized but appear as an important link between small RNA biogenesis pathways and other cells pathways (e.g. interferon pathway via PKR). At the time of writing no structure showing the interaction between two dsRBDs has been reported, even outside of the miRNAs field. The discussion is centered on the biochemical properties of these interactionsin the model organisms considered here.

4.4.1 Both dsRNA and dsRBD/dsRBD interaction capabilities within a dsRBD?

One of the most common protein/protein interaction mediated by dsRBDs in mi/siRNA biogenesis is the interaction mediated by the C-terminal dsRBD of RNaseIII cofactors. In a few case this region might not be a dsRBD. Homo and hetero dimerization have been reported for TRBP and PACT, using yeast two hybrids (42), between PACT and PKR in an RNA independent fashion (217), and with PACT itself *in vivo* and *in vitro* (218). For PACT (218) reports the region containing the first two dsRBD as involved in a homodimerization process, whereas authors of a review mention they couldn't demonstrate dimerization with the first two dsRBDs of PACT (58). Interactions between the first two dsRBDs of PACT have been suggested (220) but the authors did not rule out the possibility of an interaction mediated by an RNA. Similarly for TRBP, the role of the dsRBDs with a dsRNA binding affinity in these protein/protein interactions is not clear in these studies using mainly pull down assays (dsRBD 1 and 2). In PKR, the abilities of the first dsRBD to dimerize and of the second to oligomerize have been reported (265). The dsRBD-2 of HYL1 has less affinity than the first one for dsRNA. According to the study reporting a crystal structure of the dsRBD2 of HYL1 (294), the second dsRBD of HYL1 has a putative dimerization interface but the data provided do not enable to conclude definitively of the relevance of the potential protein/protein interaction interface in which this domain is involved. Additionally, the affinity of the proposed interaction is not known (294). Deeper biophysical studies will be required to assess the relevance of the interaction in solution and to check the binding interface suggested by the crystal structure. Additionally an independent study found that the second dsRBD of HYL1 behaves as a monomer in solution (231). However the second dsRBD of HYL1 can interacts with the central non canonical dsRBD of DCL-1 (227) with weak dsRNAs affinity. Pull down assay have reported that the second dsRBD of HYL1 is essential for HYL1 interaction with DCL-1 (122).

All these data are rather unclear and not consistent between the studies. Surprisingly, in an article using NMR spectroscopy, ITC and EMSA to underline a weak interaction between a dsRBD and a dsRNA (227), only pull down assays are used to show the existence of a dsRBD/protein interaction, a property presented a central in the article. It seems strange that these dsRBD/dsRBD interaction are almost only studied by pull down assays, a useful technique that has however many limitations in particular its inability to prove direct interactions. No results have been reported so far using techniques different from the pull down experiments, apart a complement by Far western assay with denaturing gel in one study centered on the interaction between HYL1 and DCL-1 by pull downs (122). This might reflect difficulties in handling these interacting domains or indicate very low binding affinities that can be hard to detect with other methods. However, some the structure of some of these domains (PACT-dsRBD1) have been used for structural studies that require high concentration. Thus, because of the systematic report of dsRBD/dsRBD protein interaction reported with pull down assays during the last 18 years, the relevance of the results could also be questioned. In (261) the authors mentioned that "Freshly purified full-length HYL1 is monomeric but multimerization occurred after prolonged incubation of the protein" so additional biophysical data are required to prove the relevance of the pull down assays, with emphasis on the relevance of the incubation times. The safe conclusion at this stage for dsRBD/dsRBD interaction of dsRBD with dsRNAs affinity seems to be that in pull down assays, some dsRBDs give the same result that the one expected for a protein with self associating capabilities and that such interaction have not be published with any biophysical method.

4.4.2 The C-terminal domain of dsRBP cofactors as a mediator of multiple protein/protein interactions

Despite the unclear results for the aforementioned dsRBD/dsRBD interactions, the third dsRBD of both PACT and TRBP that apparently do not bind dsRNA are consistently reported to be involved in protein/protein interactions (154). They mediate both their homodimerization and interaction with other proteins including PKR and Dicer. The third domain of PACT with only weak binding affinity for PKR is involved in the activation of PKR (220). The C-terminal region of TRBP interacts with a 165 residue region of Dicer located between the ATPase and the helicase domain of Dicer (48). Interestingly, an autoinhibition process of human dicer by its N-terminal domain including the helicase domain has been shown (178) for a perfect duplex dsRNA subtract of 35 base pair RNA with perfect base pairing. With this RNA, the full length Dicer is 65-fold less effective to process the RNA duplex than its catalytic core. The autoinhibition effects is very small with pre-miRNA tested (pre-miR-hlet-7) in (178) and pre-miR-30a (167). This differential inhibition suggests that the helicase could contribute to distinguish the hairpin substrate with the shorter duplex product. A potential role of TRBP in relation with a regulation of this inhibition related to the binding of its third dsRBD is possible. In *Arabidopsis thaliana* HYL1 can be immunoprecipitated with DCL1 via DCL1-B in the presence of RNase A (149) ; the direct interaction between these proteins has been proposed with pull down assays. A HYL1 mutant disturbs miRNAs expression whereas a mutation in the ATPase/DExH-box RNA helicase domain of DCL-1 has been reported to restore miRNAs expression (261). The results suggest a relation between the helicase domain of DCL-1 and HYL1.

Interestingly in *C. elegans*, RDE-4 can form a homodimer in solution mediated by the C-terminal domain (214) which may not be a dsRBD. This interaction increases the affinity of RDE-4 for long dsRNAs (214). The C-terminal domain is required for siRNA biogenesis (214). RDE-4 can also interact with Dicer-1, RDE-1 and DRH-1, a conserved DExH-box helicase required for RNAi in *C. elegans* (260). Interestingly, this helicase protein is the most similar protein in *C. elegans* to the helicase of Dicer-1 (260). In *Drosophila melanogaster*, the C-ter of R2D2 is required for binding to Dicer-2 (170) and R2D2 binds to the helicase domain of Dicer-2 during endo-siRNA biogenesis (212). Dicer-2 and R2D2 mutually stabilize *in vivo* (173). Similarly, the C-terminus specific part of the PD isoform of Loqs-PD can interact with the DExD/D-helicase domain of Dicer-2 (120). As with R2D2, Loqs-PD binds to the helicase domain of Dicer-2 during endo-siRNA biogenesis (212). Again, in fly, the C-terminus of these proteins might not be a dsRBD but the effects and partners of the interactions strongly resemble, suggesting an analogy within these C-terminal regions.

Structural characterization of the helicase interaction with non canonical dsRBDs and of its direct effects on Dicer awaits further characterization.

4.4.3 Simultaneous dsRNA and protein interaction in the dsRBDs of HEN-1 enzyme

A structure of *Arabidopsis thaliana* S-adenosyl-l-methionine-dependent methyltransferase Hua Enhancer 1 (HEN1), presented in the part on mi/siRNA in plants was solved by x-ray diffraction (127). The structure shows the two isolated dsRBDs interacting with a 22 nt a miRNA duplex similar to miR173/miR173*. One of the dsRBDs has a resolved unusually long β1-β2 loop involved in a protein interaction with the methyltransferase domain. This structure is particularly interesting as it is the only structure of a full dsRBD containing protein with its substrate in si/miRNA biogenesis and of a dsRBD with both dsRNAs affinity and protein/protein interaction properties. It strongly suggests that dsRBDs can help to orient other domains via both dsRNA and protein/protein interactions, including the catalytic domains around the RNA substrate.

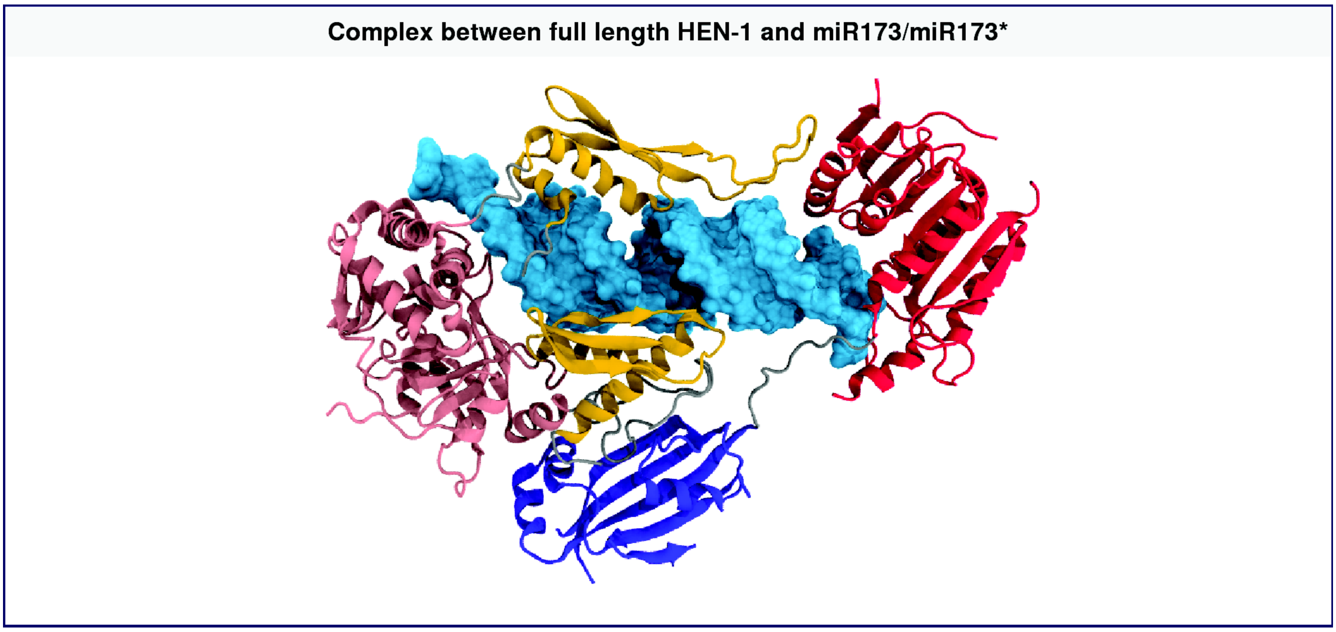

Complex between full length HEN-1 and miR173/miR173*

Figure 4.2: Structure of *Arabidopsis thaliana* HEN-1 in complex with a 22 nt duplex similar to miR173/miR173*. The dsRBDs are represented in yellow, the LCD domains in pink, the PLD domain in purple and the methyltransferase domain in red. The structure shows a dsRBD (on top) involved in both dsRNA and protein/protein interaction with the methyltransferase domain via its long β1-β2 loop.

4.4.4 Protein/protein interactions with dsRBD and subcellular localization in mi/siRNA biogenesis

Schizosaccharomyces pombe Dicer enzyme Dcr1 is not involved in small silencing RNA biogenesis but in heterochromatin assembly. Its C-terminal dsRBD has the ability to bind nucleic acid but this binding property is dispensable for the function of Dcr1 (13). This domain has however a zinc coordination site that is essential for the nuclear localization of Dcr1 and its function in heterochromatin assembly, suggesting that the main function of Dcr1 C-terminal dsRBD is nucleocytoplasmic trafficking. Interestingly, the second dsRBD of *Arabidopsis thaliana* DCL1, DCL1-B is essential for the localization of DCL-1 in the so-called dicing bodies in the nucleus (73). This is might be mediated by a protein/protein interaction in which DCL1-B is involved (149). This suggested a potential broad importance of C-terminal dsRBD of Dicer in their localization.

In human, pri-miRNAs can be edited by ADAR1 or ADAR2 (295). ADAR1 shuttles between the cytoplasm and the nucleus. The nuclear localization of ADAR1 depends on the interaction between its third dsRBD and transportin-1 (TRN1) (90). The third dsRBD of ADAR1 appear highly homologous to other dsRBD. DsRNA binding by ADAR1, involving the third dsRBD abolish nuclear import via TRN1 binding and its export via exportin-5 binding (90). Thus ADAR dsRBDs appear involved in the signalization of a dsRNA sensitive nucleocytoplasmic transport.

Despite human Dicer and TRBP do not have known nuclear localization signals and are mainly present in the cytoplasm, some studies have found that a fraction of these proteins localize in the nucleoplasm (7). Human Dicer localization in the nucleoplasm seems to be mediated via interaction with nucleoporins, co-immunoprecipitated with Dicer (7).

Conclusion Thus, dsRBDs are a particularly interesting pleiotropic small protein domain in mi/siRNA biogenesis despite globally similar folds. Many roles and original phenomenon are frequently discovered like the recent diffusion property on dsRNAs and on the contrary some are getting less reported like the annealing activity of dsRBD. It was demonstrated that the dsRBD of Xlrbpa promotes RNA strand annealing i.e. RNA chaperone activity (123), not correlated with their affinity for dsRNAs. This activity is found in the full length protein and also with its isolated dsRBDs. A similar annealing activity has been found in PKR (265), which seems to be related to a straightening effect of PKR dsRBDs on a RNA/DNA hybrid with a bulge (305). This could be linked to a cooperativity between the dsRBD, the binding of one could help the other dsRBD of a tandem or from different proteins. However there is lack of structural information of dsRBD with imperfect A helix dsRNA so the significance and extension of this process with different dsRBDs and dsRNAs remains to be determined. This is in fact a general comment of the data available on dsRBDs in metazoan miRNA biogenesis. Despite almost all metazoan pre-miRNAs are imperfect dsRNA structures, the relation of these irregular features with the binding of dsRBDs remains very elusive. In the same spirit, the mutual effects of the dsRBDs interactions with miRNA precursors are still very unclear.

Part II

General strategy and biophysical techniques

This part starts by a chapter that explicit the scientific strategy that frames all the work I report here. This strategy has been applied through the usage of biophysical techniques that are the object of the following chapters.

This part contains the following chapters :

- **Chapter 5**: Strategy: system, workflow and biophysical techniques
- **Chapter 6**: Characterization of chemical equilibriums in solution
- **Chapter 7**: Principles, implementation and interpretation of Size Exclusion Chromatography (SEC)
- **Chapter 8**: Principles, implementation and interpretation of Size Exclusion Chromatography coupled to Multiple Angle Laser Light Scattering (SEC-MALLS)
- **Chapter 9**: Principles, implementation and interpretation of Sodium Dodecyl Sulfate Polyacrylamide gel electrophoresis (SDS-PAGE)
- **Chapter 10**: Principles, implementation and interpretation of isothermal titration calorimetry for the study of chemical equilibrium
- **Chapter 11**: Principles, implementation and interpretation of sedimentation velocity Analytical Ultra-Centrifugation for the study of reversible interactions.
- **Chapter 12**: Introduction to NMR spectroscopy to study secondary structures and some exchange processes

It is not the goal of this part to develop all the theoretical and practical aspects of the biophysical methods used. This practice would have contaminated and diluted the thesis work reported here. Excellent books and articles already exist for most of the biophysical methods used and are given as references. However, together with the description of the technique implementation used in this work, it was necessary for each technique to present the key properties required to understand the origin of the signal, the assumptions used to derive the theoretical framework of the technique, the biases and the non trivial processes leading to indistinguishable results. This is necessary to understand the data analysis and the exact contribution of the experiments to the knowledge of the system. The next parts of this thesis will strongly be based on this part. The motivations that led progressively to the choice of specific techniques are developed in Chapter 5.

The Nuclear Magnetic Resonance (NMR) spectroscopy chapter is not as developed as the other methods taking into account the large theory and diversity of the NMR experiments. A lot as many excellent books are devoted to this and many theses written in the host laboratory have already detailed the NMR methods developed here and used. The main NMR theoretical and practical aspects as well as the complementarity with the other methods are presented.

Chapter 5

Strategy: system, workflow and biophysical techniques

5.1 Definition of the biological systems to be studied

5.1.1 RNA binding region of TRBP

The ScanProsite server (50) was used on the sequence of the long isoform of TRBP (TRBP1) to get a preliminary prediction of the boundaries of the dsRBDs of TRBP. The program predicted 3 dsRBDs in positions 30-97, 159-227 and 298-361. The available structures of TRBP dsRBD1 and dsRBD2 solved before the start of the thesis were perfectly compatible with this prediction using the residues 22-105 and 150-225 respectively. To complement this domain prediction, a secondary structure prediction and a prediction of disordered segments was performed by the two open source programs Psipred (133; 191) and Disopred2 (276) respectively. Psipred predicts 3 states for each residue (helical, sheet, coil), the percentage of the ones being correct (Q3 score) is estimated to be 77 %. Disopred2 is based on a database of missing fragments in solved crystal structure, a simple approach recognized as biased by the authors of the algorithm because of the multiple sources of missing fragments. This was used to define potential addition of a few residues at the boundaries of the predicted domains to avoid cutting potentially longer secondary structural motifs. The last dsRBD of TRBP does not interacts with dsRNAs (147) and since this dsRBD and potentially the interdomain region between dsRBD2 and dsRBD3 are predicted to be involved in protein/protein interactions and not in the interactions with RNA, this work has focused on the RNA binding region of TRBP comprising the first two dsRBDs. Attempts in this thesis to work at high concentration with full length TRBP confirmed, as explicited in some articles, the difficulty to work with this protein when isolated. Protein constructs have been designed before the start of the thesis in order to analyze the RNA binding region of TRBP. The idea was to study the full RNA binding region as well as isolated domains to compare their behavior in isolated form with their behavior in the RNA binding region. This led to the following constructs with

TRBP meaning the TRBP2 isoform and TRBP1 the elongated isoform that has a 21 more residues at its N-terminus: TRBP-dsRBD1 (19-99), TRBP1-dsRBD1 (1-99) TRBP-dsRBD2 (156-228), RNA binding region of TRBP, TRBP-dsRBD1&2 (19-228, including the two first dsRBDs). The name of the TRBP isoform will be only used for the construct TRBP1-dsRBD1 (1-99) including the N-terminal extension corresponding to the isoform TRBP1. Inclusion of 3 residues from TRBP1 in the TRBP1-dsRBD1 (1-99) construct enable an easier NMR assignment of the first residues of TRBP2, starting at position 22. The sequences and boundaries of the constructs TRBP1-dsRBD1 (1-99) and TRBP-dsRBD2 (156-228) are given in **fig. 5.1** .

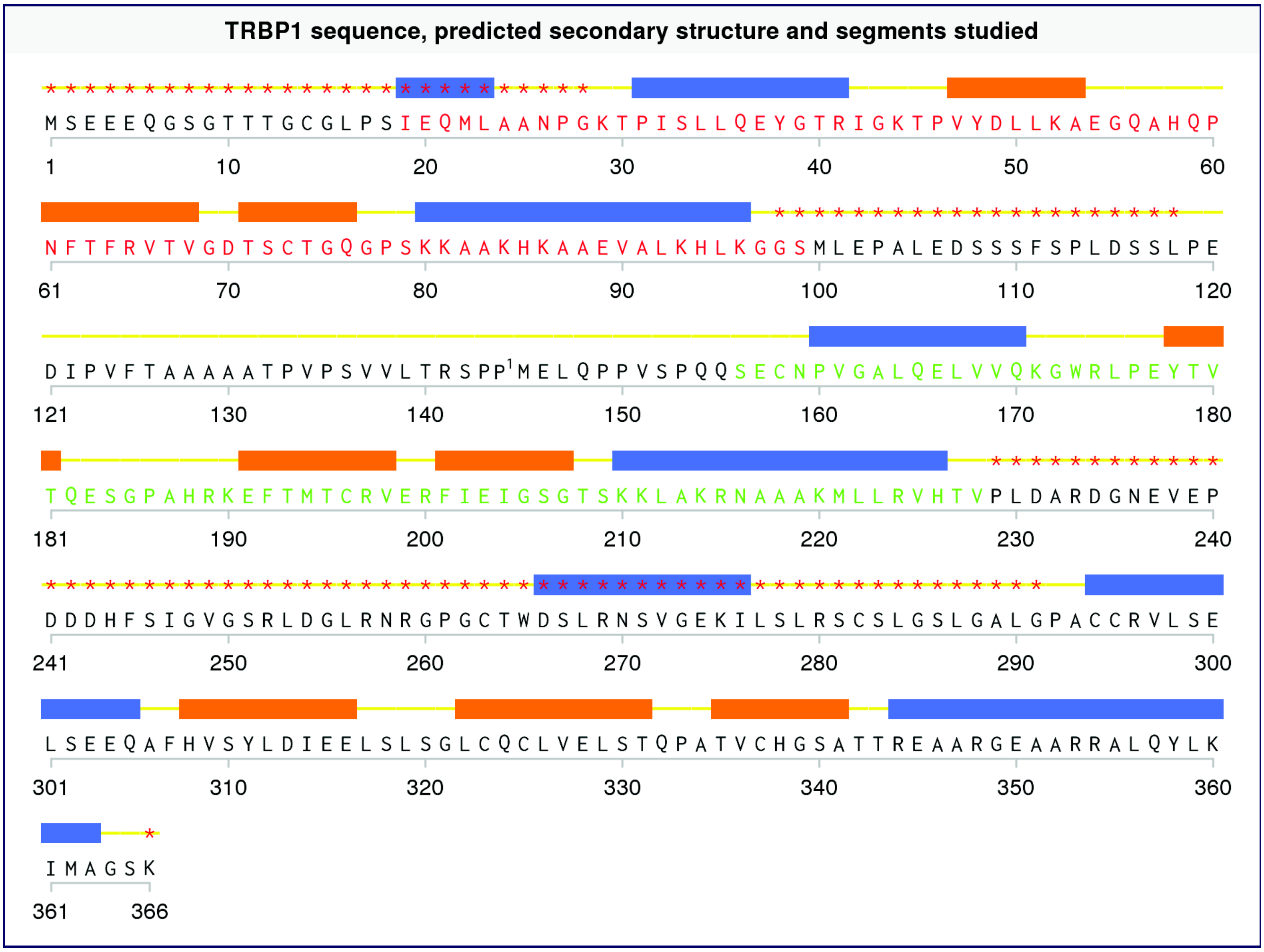


Figure 5.1: Sequence of TRBP1 (uniprot Q15633) and coloration of residues of TRBP-dsRBD1 (red) and TRBP-dsRBD2 (green) constructs. 1: proline 144 reported as an alanine in the original sequence of TRBP ((97), GeneBank: M60801). TRBP2 isoform starts at residue 22. The PSIPRED secondary structure prediction are indicated above the sequence using orange for sheets, blue for helices and yellow for coils. Disopred2 prediction of disordered regions is indicated by red stars "*".

The TRBP sequence used for the designed work was performed with the original reported TRBP sequence by A. Gatignol (97) in 1991. The constructs were designed in December 2007. In October 2008, a sequence conflict was reported at position 144 in the interdomain region between the first two dsRBDs because of a mis-identified C nucleotide replaced by a G in the original sequence. As a result in the protein sequence at position 144, the reported alanine is actually a proline (139-TRSPA-144 protein motif instead of 139-TRSPP-144). This is indeed an error since a tblastn (5) performed with TRBP1 residues 18 to 228 against the NCBI human Expressed Sequence Tags (EST) database and the NCBI human nucleotide sequences database on April 2013 did not revealed the protein motif "TRSPA" but only "TRSPP" apart from the first reported TRBP sequence. The proline to alanine mutation has not been reported to date in any article devoted to TRBP even in the successive articles of Gatignol on TRBP including a TRBP review in 2013 (47) and is still not mentioned in the original GeneBank entry (M60801, with its only version M60801.1). The actual TRBP sequence used in most articles is not clear. The original sequence with the alanine 144 has been used in this project. The finding in this thesis of the very high flexibility of the interdomain region of TRBP make that no changes are expected in relation to this mutation for the global properties studied here regarding the free protein and its interaction with miR-155 precursors. A slight rigidification localized to the residues 143 and 144 in the sequence, forming a proline-proline motif is possible, similar to the one seen in the other proline-proline motif of the linker. A natural serine/phenylalanine variant of TRBP at position 251 corresponding to the interdomain region of TRBP has also been reported. This variant belongs to a TRBP region which is not part of the RNA binding region construct used in this thesis.

5.1.2 Precursors of hsa-miR-155

The two precursors of hsa-miR-155, corresponding to the steps before and after Dicer cleavage and expected to interact with TRBP were selected in this study fig. 5.2 .

A. **pre-hsa-miR-155, before Dicer cleavage**

```
                  c   a        u-   cc
   5'P-uuaaugcuaau gug uagggguu  uug  \
3'OH-acaauuacgauua-uac-auccucag  aac  /
                               uc   cu
```

B. **hsa-miR-155-5p / hsa-miR-155-3p duplex**

```
                  c   a
   5'P-uuaaugcuaau gug uaggggu-3'OH         hsa-miR-155-5p
3'OH-acaauuacgauua-uac-auccuc-5'P           hsa-miR-155-3p (*)
```

Figure 5.2: Predicted secondary structure of two selected successive precursors of the human oncogenic microRNA hsa-miR-155-5p, before and after Dicer cleavage. **A.** Secondary structure of the pre-miRNA 155 after cleavage by Drosha/DGCR8 (microprocessor) complex. **B.** Secondary structure of the miRNA/miRNA* duplex of microRNA 155 after Dicer cleavage.

5.2 Multiple working hypotheses and associated technique presentation

Multiple working hypotheses Experimental methods that give results that can unambiguously be linked to a single phenomenon are very rare, if they exist. This motivates the multiple working hypotheses work flow firstly clearly reported by the geologist Chamberlin in 1890 in the old series of *Science* (32), reprinted in (33). The approach used throughout this work is consistent with this work flow. Instead of focusing on a leading (exiting) hypothesis and checking its consistency with the results, the multiple working hypotheses consists of the definition of all the processes parallel or successive, compatible or exclusive, that could lead to the result observed given the current knowledge of the experimental system. Additional experiments or information can then restrain the number of these working hypotheses. Chamberlin article has lately been followed by various articles on "strong inference" a concept named by Pratt (222). Although interesting and being the source of a flow of informative scientific and philosophical papers, strong inference is only a subset of the Chamberlin strategy in which multiples hypotheses are not obviously exclusive. A recent article has aimed to underline these differences (69). Multiple working hypotheses strategy has found recent quantitative analogy with the development of information theoretic approaches including the Akaike Information Criterion and Akaike weights for model selection. This quantitative approach will also be used in this work. Distinct mechanisms that can lead to the same apparent result are very scarcely reported in the texts devoted to the teaching of biophysical techniques. They will be particularly underlined in the second part of this thesis since they are essential for an appropriate scientific reasoning with the available results. These principles motivated the way the experimental methods are presented, as developed in the next paragraph.

Structure of the biophysical techniques presentation

- *Principles and implementation*, explains the origin of the signal observed and its relations with some properties of the system under study. The exact implementation of the technique in this work is given here including the material and devices used. The juxtaposition between the principles and the implementation in two successive paragraphs rather than in different chapters aims to have a compact document and so that the protocols steps justification are easily understandable for a wide range of readers with different backgrounds.
- *Data analysis and related approximations and biases*, gives the theory used for the data analysis used in this work and insists on known systematic biases and on the multiple approximations that defines condition to be checked during a data analysis.
- *Non-trivial indiscernible processes*. Among all the processes that can contribute to the signal observed, whose description is the goal of the "Principles and implementation" parts, some combination will lead to the same signal. This can correspond either to two similar processes in the system with only one being detected, either to two processes that generate indistinguishable signals. These combinations of processes are then indistinguishable without additional dedicated

experiments. I define here what I called "non-trivial indiscernible processes", as indistinguishable process that can lead to the same result for a given experimental technique. Trivial sources of confusion will not be developed in this part, like the fact that the set of possible results is usually constrained by the experimental device. The "non-trivial indiscernible processes" could also have been called sources of confusions, sources of indistinguishability, sources of ambiguities, sources of multiple interpretations. It is my point of view that the presentation of experimental data analysis and associated conclusions cannot be presented safely without the addition in the conclusion of the list of the main processes that can lead to similar experimental results. For a clear identification, the non trivial indiscernible processes defined in this step of the presentation will be marked with brackets ({ }).

5.3 Selection of biophysical methods

5.3.1 Multiple compatible biophysical methods

A particular effort has been used in this thesis to use, when possible and reasonable, various experimental in solution biophysical methods to study the system. Each biophysical technique is sensitive to a subset of properties of the system under study and has a set of biases. As a consequence - already underlined - almost systematically, multiple hypotheses on the system can explain a given result obtained with one method. However, the set of sensitive properties and biases of one method is usually not fully shared by another one. The combination of techniques sensitive to different properties can restrain the number of theories that could explain the set of available results. This enables to get more reliable conclusion than a single method study. The selected methods were chosen to be compatible: all had to be applicable on the same system in solution with the same physico-chemical environment (buffer, temperature, pH,etc). There is limited interest to choose different methods to build a model if we know at the beginning that the conditions are so different that highly different behavior is expected in each method.

The main in solution biophysical method used in this work, which is one of the most informative and the specialty of the lab where this work was performed, is liquid state Nuclear Magnetic Resonance (NMR) spectroscopy. It was used for the characterization of free TRBP proteins constructs, miR-155 precursors and their interaction surfaces.

5.3.2 Comparison of common methods to characterize macromolecular interactions

Selection criteria for a technique to study TRBP/miR-155 macroscopic equilibrium In order to define the macroscopic equilibrium (stoichiometries and binding constants) between TRBP proteins constructs and miR-155 precursors, a selection of additional methods has been performed among the ones enabling to do this.

Table 5.1: Comparison of methods to characterize macroscopic equilibriums *in vitro*. In the columns "dissociation constant estimation" and "stoichiometry estimation", "yes" means that is expected to be determinable in most cases but not guarantied and 1:1 means restricted to 1:1 interactions ; "possible" means that it is only reasonably determinable in a small range of situations not detailed here.

method	**abbreviation**	**molecular labeling free**	**immobilization free**	**porous matrix free**	**dissociation constant estimation**	**stoichiometry estimation**	**theoretical foundation**
Nuclear magnetic resonance	NMR	yes	yes	yes	possible	possible	quantum mechanics
Isothermal titration calorimetry	ITC	yes	yes	yes	yes	molar ratio	thermodynamics
Analytical ultra centrifugation	svAUC	yes	yes	yes	yes	yes	thermo/ hydrodynamics
Surface plasmon resonance	SPR	yes	no	yes	yes, 1:1	no	quantum mechanics
Gel shift assay	GSA	yes	yes	no	yes	yes	empirical
Cg multi-angle laser light scattering	cg-MALLS	yes	yes	yes	possible	possible	quantum mechanics
Filter binding assay	FBA	yes	no	no	yes, 1:1	no	empirical
Micro thermophoresis	MTA	no	yes	yes	yes	no	under debate
Fluorescence polarization	FP	no	yes	yes	yes	possible	empirical

The methods that were preferred for this study were selected, to reduce biases, regarding the following criteria :

- No molecular labeling requirement for the macromolecules. In this case the experiment reports the binding properties of a chimera, and the best way to show that the chimera has no effect is to do the study without the molecular label. This appeared particularly unsuitable for the rather small macromolecules used in this study.
- No binding requirement for the protein or RNA on a surface or interaction in a solid (*e.g.* filter binding assay, surface plasmon resonance). Since the system under study is supposed to be soluble, this binding can alter the availability of the binding sites. The dynamics of the system can also be affected and the dynamics is well known to be have links with binding properties.
- Homogeneous technique *i.e.* that does not require separation of bound and free forms that could alter the equilibrium or the stoichiometry, like a porous matrix or a crystal.

5.3.3 Selection of methods for binding constant estimation and stoichiometry determination

The goal here is not to review the following methods but to present the mains strengths and weaknesses of the methods regarding the previously criteria adapted for the of this thesis **tab. 5.1**. Additionally to the labeling requirement, the size of TRBP[19-228] (22 kDa) and of miR-155 precursors (> 15 kDa) made not fluorescence polarization a good choice since it is suitable for ligands of molecular weight under 10 kDa (2). Thus, among the methods compared and with the selected criteria, NMR, ITC and AUC appear to be good candidates for the characterization of the macroscopic equilibrium of TRBP and microRNA precursors. In well behaved systems for NMR, this method can be used to estimate the dissociation constants for all detected resolved sites. However, in many cases where intermediate exchange happen, some observed nucleus have severe line broadening that strongly complicate this analysis and it can be hard to get the stoechiometry, which is a key parameter. Additionally with the standard labeling techniques some large complexes can be missed in the analysis. For reactions with non zero reaction enthalpies, ITC can often give fast determinations of dissociation constants and molar ratios in a more direct way than NMR and AUC. It can also provide estimation of binding enthalpies. Thus ITC has been selected for the macroscopic equilibrium characterization. However, as underlined in **chapter 10**, ITC does not give the exact stoichiometries but molar ratios, so the validity of the binding constants requires a check of the actual stoichiometry. Additionally some reactions with almost zero enthalpy of reaction do not give ITC signal and can be missed by ITC. It was thus required to complement ITC with another method to study the stoichiometry. Ideally the method used to confirm the molar ratio and to determine the stoichiometry should not be based on the fitting of a one dimensional curve by increasingly complicated models, like what was done in ITC. A more direct evidence for the stoichiometry was preferred.

Motivations for the choice of ultracentrifugation methods for stoichiometry determination of the interaction Although an apparently simple parameter, potential stoichiometry values of a system are usually difficult to constrain experimentally. Few methods can do this rather directly at the time of writing. The main ones are given in the following list with their respective main advantages and limitation :

- Native mass spectrometry is sensitive to the mass of the complex in a gas phase. It has not constraints on the molecular weight and can enable to propose combination of the masses of the individual partners that match the mass of the complex. This can lead to constrained and often unique possibilities of stoichiometries. However this method does not work in liquid state, is restricted by now to stable protein complexes and not yet mature for transient protein/RNA interactions.
- Cross linking strategies have been proposed for protein/RNA complexes (116). The method is based on an irradiation with ultraviolet light (254 nm) that triggers the cross linking between nucleotides and residues. It does not require chemical modification of either the proteins or the RNA. Then a "native" mass spectrometry analysis can be used to get information about the stoichiometry. Although elegant, there are no guaranties that the cross linkers will only cross linked relevant interactions or inducing structural changes that might impact the network of interactions.
- Size exclusion chromatography and MALLS. SEC is not perfectly in solution, involve separation of free and bound species that can affect the equilibrium and as already mentioned, has no rigorous theoretical foundation to analyze the data when an interaction happen. It could be coupled to MALLS but it seems very hard to satisfy the classical MALLS assumption (see chapter 8) with a transient complex. Additionally if a complex elute in one peak in SEC, MALLS data analysis will give a weight average of its components that will be difficult to interpret.
- Composition-Gradient Multi-Angle static Light Scattering (CG-MALS) is based on the measurement of the light scattering signal from a series of solutions of different macromolecular composition. This method is truly in solution, has foundation in thermodynamics and enables to record data in a few minutes. However, it does not allow to get directly a distribution of the species in solution. Data analysis requires the additional methods like SEC-MALLS which can be ambiguous with transient complexes. Finally, this powerful method did not appear to be the best choice after the ITC analysis since it also requires fitting a one dimensional curve with models of varying stoichiometry.
- Sedimentation study by ultracentrifugation works in solution for any type of macromolecule and for a wide range of size up to 100 kDa. The method has foundations in thermodynamics and hydrodynamics ; all the parameters of the equations that describe sedimentation behavior are experimentally determinable. It however requires long experiments of at least 5 to 6 hours. It enables to get many curves that can be processed to get information on the size/sedimentation coefficient distribution of species in solution. Rather direct information of the stoichiometry can

then be extracted from these distributions. Because of the large differences in its theoretical foundations compared to ITC and because of the possibility of getting information on the distribution of species in solution, this method has been selected to complement the ITC data. Moreover, a wide range of dissociation constants can be studied in AUC (dissociation constants between 10^{-9} and 10^{-4}). The two dimensional HN NMR experiments and one dimensional proton experiments have shown that the protein and RNA samples used in this work were stable for longer period than the one required for svAUC.

Comparison of AUC experimental methods and motivations for the use of sedimentation velocity experiments The idea of the ultracentrifugation techniques is to get hydrodynamic parameters from the study of the sedimentation of macromolecules in a strong gravitational field of over 10000 g. There are two complementary methods to study biological macromolecules with an ultracentrifuge: sedimentation velocity (svAUC) and sedimentation equilibrium experiments. Sedimentation velocity uses a strong gravitational field and long cells to maximize the resolution. It can give information about size, shape, masses and molecular interactions. Sedimentation equilibrium uses a lower gravitational field and short cells in order to reach faster the equilibrium. It can provide information about molecular mass, association constants and stoichiometry. Both methods can be used to study interactions. However, the resolution of sedimentation equilibrium is lower and the required long experimental time, a few days, is a major constraint for many samples compared to the shorter one for sedimentation velocity (4-6 hours). The former method also requires very high quality data that are not guaranteed *a priori*. Since the AUC method was mainly used to confirm the molar ratio proposed in the ITC analysis and to determine the stoichiometry, to avoid the potential apparition of irrelevant species due to the long storage of the particle, sedimentation velocity experiments were used here.

5.3.4 Main biophysical methods used

Thus NMR studies were associated with other in solution methods: Isothermal Titration Calorimetry (ITC), Sedimentation Velocity Analytical Ultra Centrifugation (svAUC) and Multiple Angle Laser Light Scattering (MALLS). These methods were complemented by size Exclusion Chromatography (SEC) and SDS-Polyacrylamide gel electrophoresis. Although not purely in solution techniques, the last two methods were useful for their contribution to the purification process and to the purity analysis. Each method gave some information on the system that couldn't be safely extracted from the other methods used. They also gave information that can be used to check that there is some common behavior between the techniques. All this information has then been assembled to build a model of the interaction between the RNA binding region of TRBP and the stem-loop and duplex precursors of the human microRNA 155.

Chapter 6

Characterization of chemical equilibriums in solution

6.1 Macroscopic and microscopic chemical equilibriums

Non covalent interactions between molecules involve weak forces including electrostatic and hydrogen bounding forces. Also called transient interactions, they are involved in most life processes. Their characterization is one of the key to understand biological processes and they are also of great interest for drug design. An atomic resolution knowledge of inter macromolecular interactions would require to know the magnitudes of these forces or potentials for the set of dynamically interacting atoms. Another simpler and useful way to characterize the interaction is to consider sets of interacting atoms. Two description of these set are usually reported: we can either differentiate each sets of sites that interacts ("microscopic" description of the equilibrium) either only differentiate the different sets of species that can associate ("macroscopic" description of an equilibrium). Each equilibrium can be defined by a chemical equation, representing mass and charge conservation as illustrated in **fig. 6.1**.

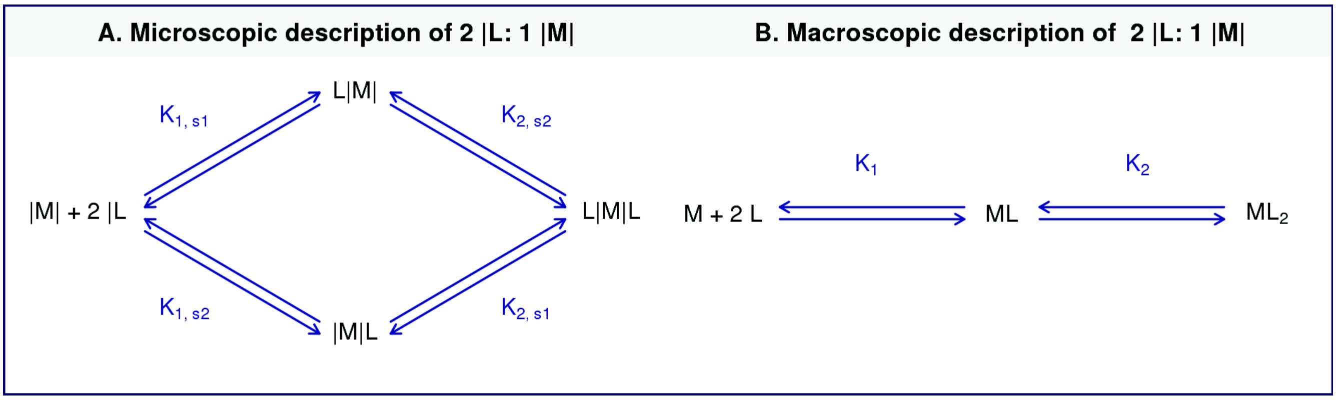


Figure 6.1: Illustration of the microscopic description (A.) and the microscopic description (B.) of a chemical equilibrium involving two species M and L. M has two interaction sites (s_1 and s_2) for L and L has one interacting site for M. The sites are represented with vertical bars (|). Microscopic and macroscopic association constants names are given for each sub equilibrium and defined later in the main text.

Both microscopic and macroscopic equations of the equilibriums can be quantitatively complemented with thermodynamic and kinetic parameters. In particular the study of the system at equilibrium enabless to get estimates of the concentrations of the species at equilibrium and so quantify how displaced is the equilibrium and so the strength of the interaction. The full microscopic description is richer in information than the macroscopic one since the exchange between each assembly state can involve multiple paths according to the sites involved **fig. 6.1**. An interest of the macroscopic interaction is that it is sufficient to globally compare the behavior of various molecules and give very useful information to design a study at a higher precision. The goal of this part is to present the main classical thermodynamic theory for the study of the macroscopic equilibrium from titration experiments. The development will focus of the necessary approximations and the relation between the quantity actually measured and the quantities of real interest.

6.2 Thermodynamical quantification of equilibrium displacement

6.2.1 System, environment and strategy to study macromolecular interactions

Interacting molecules The system in this study contains the protein TRBP and microRNA precursors, often mixed together to study their potential interaction. If an interaction happens, the interacting system might not be reduced to different associations of these macromolecules but additional species from the surrounding media can participate in the reaction. These additional species can be protons, metal ions that neutralize the charged residues, etc.

Overall strategy The common general strategy to study equilibrium is to monitor during a titration some physico-chemical probes perturbed upon the interactions as a function of variables representing the amount of species introduced progressively in the system. This is what have been done is this work with different methods. The deconvolution of the sources of perturbation, not only reduced to the interaction process, is the key to the quantification of the interaction. Specific development will be devoted to this analysis in the chapters devoted to the biophysical techniques used for this study. Two phenomena common to each technique are worth to define here and consist the first two approximations in the following development :

- *Approximation* $\mathcal{A}_{Inter}0$: biologically reasonable interaction conditions. In the interaction characterization of macromolecules, some interactions contribution are qualified specific while others are said to be nonspecific. The original idea behind this distinction seemed to separate the relevant versus irrelevant interactions. Although commonly used, these terms are hard to define since the two situations are described by the same physical forces. An extreme case to illustrate the idea of irrelevant interactions is the aggregation of some monomeric proteins *in vitro* environment with very low salts concentration, whereas no such protein/protein interactions exists *in vivo*. Interesting discussion on the meaning of "specific" *vs.* "nonspecific" interactions have already been published in many places but without improvements of the concepts, for instance in (11). The point of view adopted in this work was to choose chemical environment conditions for the *in vitro* studies not too far from the *in vivo* ones in temperature, pressure, ionic strength and pH. These simplified conditions aim to be able to underline the relevant fundamental properties of the isolated molecules, despite the intracellular effects (crowding, etc) are not taken into account. This is one of the key assumptions for *in vitro* studies, enabling to discuss the results in the lights of the *in vivo* data available. If the assumption is not fulfilled, the study is reduced to the biophysical study of the interaction of two biological polymers.

- *Approximation* $\mathcal{A}_{Inter}1$: negligible macromolecular alteration. The first approximation in the interaction studies is that potential temporal physico-chemical variation of the macromolecules between the initial and the final state have to be negligible compared to the changes linked to the interactions. These undesirable physico-chemical changes can be caused by oxidation reactions (so the importance of a reducing agent for non disulfite bound containing proteins), enzymatically catalyzed reactions like peptide bound cleavage that can be caused for instance by a bacterial contamination. The latter phenomenon should be negligible for highly pure macromolecules and will be checked for the different methods. This assumption is referred as $\mathcal{A}_{Inter}1$.

6.2.2 Macroscopic equilibrium constant *vs* measured concentration constant

The following development contains important assumptions that are required to derive the standard relations used in the data analysis for the thermodynamical characterization of the equilibrium. These assumptions are commonly not reported in the biochemical studies and while their explicitation is important to modulate the conclusions obtained with methods used to characterize equilibrium constants.

Macroscopic equilibrium constant Let's consider a closed system in which happen the following reversible reaction, with n species ($n \in \mathbb{N}^*$) $A_{j,j\in\{1,n\}}$ of stoichoimetric coefficients $\nu_{j,j\in\{1,n\}}$:

$$\sum_{j=1}^{n} \nu_j A_j = 0 \tag{6.1}$$

The equilibrium will be studied at constant exterior pressure P and temperature T in this work with variable molar quantities of the species involved $n_{j,j\in[\![1,n]\!]}$. In these conditions, the appropriate energy potential is the Gibbs energy $G(T, P, n_{j,j\in[\![1,n]\!]})$. Using the fact Gibbs energy is a state function of the variables pressure P, temperature T and of the molar quantities of the n species j, N_j and using the second principle of thermodynamics, the equilibrium condition is :

$$\sum_{j=1}^{n} \nu_j \underbrace{\left(\frac{\partial G}{\partial n_j}\right)}_{\mu_j} = 0 \tag{6.2}$$

with the $\mu_{j,j\in[\![1,n]\!]}$ term called the chemical potential of the species j. By similarity to the expression of the chemical potential of an ideal gas, the composition dependency $(x_{j,j\in\{1,n\}})$ of the thermodynamic potential of a type of solutions with N components is expressed with a reference chemical potential $\mu_j^{ref}(T,P)$ as :

$$\forall\, j \in [\![1,n]\!],\ \mu_j = \mu_j^{ref}(T,P) + RT\ln(x_j) \tag{6.3}$$

This relation defines what is called an ideal solution (53). Some consequences of this definition is that for any pair of molecule in the solution, the mean strength of the interactions are the same, the enthalpy of mixing of the reactants and products mixed in their equilibrium proportions is zero and the volumes are additive. However and importantly, this equation does not describe real solutions for which the two previous implications are usually not verified experimentally. An additional term in the logarithm has thus been added (53) to account for this behavior and misleadingly called activity "coefficient". This coefficient, noted γ_j is actually a function of T, P and the molar fraction of all the species in solution. Thus the general expression of the variation of chemical potential of the species j at constant P and T in a real solution can be expressed as :

$$\forall\, j \in [\![1,n]\!],\ \mu_j = \mu_j^{ref}(T,P) + RT\ln(\underbrace{\gamma_j(T,P,x_{1:N})x_j}_{a_j}) \tag{6.4}$$

The activity of a component in solution $a_{j,j\in\{1,n\}}$ is defined as the function in the previous logarithm. The arbitrary reference standard states pressure has been chosen following UIPAC recommendation: $p^\circ = 10^5 Pa$, which is also the ambient temperature used in this study. The reference temperature is added to define this state and set to be 298.15 K. It is the temperature used in this study as recommended (8) partly because this common temperature facilitates comparisons between different studies but more biochemical details on the choice of the physico-chemical conditions of the system are developed in the part devoted to the set up of macromolecules production protocol. Since the solutes in this study are irrelevant ionic solids in their pure form in the defined standard state, following UIPAC recommendations the asymmetric solvent/solute convention is used to define the reference states of the chemical species (53), using an infinite dilution state for the reference chemical potential.

Approximation $\mathcal{A}_{Inter}2$: if the solutes participating in the reaction are very dilute, it is possible to neglect the moles of these solutes compared to the moles of the solvent. This approximation will lead a negligible bias in this study since the concentration of pure water is 55.56 M, much higher than the concentration of the solutes with a total concentration less than 200 mM. Thus, the reference chemical potential for a solute $A_{j\in\{1,n\}}$ is defined from a hypothetical state at an infinite dilution at the same pressure and temperature as the one in the study. The chemical potential expressed in concentrations using a reference concentration (C°) of 1 mol/L can be expressed as :

$$\forall j \in [\![1,n]\!],\ \mu_j(T,P) \underset{\mathcal{A}_{Inter}2}{\approx} \mu_j^{inf}(T,P) + RT\ln\left(\gamma_j(T,P,[A_{1:N}])\frac{[A_j]}{C^\circ}\right) \tag{6.5}$$

The expression of the activity coefficient is obviously the main difficulty in the use of this relation. Because of the definition of an ideal solution, the activity coefficient will approach unity as the concentration become infinitely low. Many works have aimed to give explicit expression of the activity coefficients. The related expressions are however hard to include in the modeling of the processes that can be monitored with current biophysical methods. The activity coefficient is particularly important in electrolyte solutions like in the system under study since for two ions of opposed charge z_+ and z_-, the attractive electrostatic force is inversely proportional to (only) the square distance r between these two ions: $\frac{z_+z_-}{r^2}$. For example, even H_2SO_4 a component with few charges has an activity coefficient significantly different from 1 at 500 μM in water at 25°C: 0.885 (53). Introduction of the last expression of the chemical potential into the equilibrium equation leads to :

$$\underbrace{\prod_{j=1}^{n}\left(\frac{[j]}{C^\circ}\right)^{\nu_j}}_{K_c} \underset{\mathcal{A}_{Inter}2}{\approx} \prod_{j=1}^{n}\left[\gamma_j(T,P,[A_{1:N}])\right]^{-\nu_j} \underbrace{\exp\left(-\frac{\sum_{j=1}^{n}\nu_j\mu_j^{\text{inf}}(T,P)}{RT}\right)}_{K(P,T)} \tag{6.6}$$

with the $K(T,P)$ term called the reaction constant. Practically, without any knowledge on the activity coefficient, only the concentration fraction of the equilibrium constant can be studied from titration experiments, which is called concentration "equilibrium constant" (K_c). The concentration "equilibrium constant", K_c is rigorously not a constant at fixed pressure and temperature since the activity coefficients depends on the composition of the system. The previous formula relate this concentration "constant" to the actual equilibrium constant. The closer to ideality the solution is the closer the two constant.

Another complication comes from the fact only a subset of the n reacting species might be known. For intance some species different to the biological polymers can be involved in the reaction: protons, mineral ions etc (8). These potential p species not taken into account imply to rewrite the measured concentration constant K_{cm} as :

$$K_{cm} \underset{\text{def}}{=} \prod_{j=1}^{p}\left(\frac{[j]}{C^\circ}\right)^{\nu_j} \underset{\mathcal{A}_{Inter}2}{\approx} \prod_{j=p+1}^{n}\left(\gamma_j(T,P,x_{1:N})\right)^{-\nu_j}\prod_{j=1}^{p}[j]^{-\nu_j}\ K(T,P) \tag{6.7}$$

Relevance and goals of the macroscopic equilibrium characterization in this study As a result of the previous considerations, the measured K_{cm} "concentration binding constant" is actually not exactly the binding constant K because of the approximation $\mathcal{A}_{Inter}2$ (of negligible effects), of the non ideality of the solution (i.e. activity coefficient different from 1) and of the species involved in the equilibrium and not considered in the expression of K. The last two aspects might not be negligible at all; they are not required approximations and will be mentioned in the data analysis. Furthermore, the measured concentration constant $K_{cm,\text{measured}}$ is not the exact K_{cm} for three additional reasons depending on the experimental method: potential unknown co-reactions happening in parallel to the reaction under study and contributing to changes of the concentrations of the species involved in the equilibriums, systematic biases and random errors in the estimation of K_{cm} coming from the measurement process. The relevancy of an approach neglecting the activity coefficient difference to unity or the contribution of co-reactions can be questioned without a specific studies or strong theoretical arguments. For instance, despite the macromolecules used in this study for interaction characterization are all used at low concentration (20 μM to 800 μM), they are highly charged so it does not seems reasonable to assume that the activity coefficients are equal to one. Instead of making many unrealistic assumptions, the point of view adopted in this work was to fully explicit the quantities actually measured and then discuss the meaning of the measured value in the light of the available information on the system. This enables to clearly separate the rigorous theoretical expression of the measured quantity from its interpretation and to avoid silent error propagation. Moreover, the fact the concentration constant is not the actual equilibrium constant is not a limit to achieve some characterization goals of the system. If the goal of the experiment is the estimation of the reaction Gibbs energy variation, an accurate binding constant estimation is required but it may not be the case for some applications like finding minimal conditions to shift the equilibrium towards one complex. The goal of the macroscopic equilibrium study of this work is not to have absolutes parameter estimation or to infer what happen into the cell conditions. It is mainly done, with the previous limitations in mind, to :

- define the nature of the equilibriums involved in the system
- compare with the same framework the behavior of macromolecules with different partners
- define estimates of the equilibrium constants that enable to find the conditions under which the system is saturated or under saturated, in order to design deeper biophysical investigations, including the definition of the interaction surfaces.

Chapter 7

Principles, implementation and interpretation of SEC

Definition and main biophysical applications of SEC Size exclusion chromatography (SEC) is a chromatography technique whose separation principle is based on the size of the molecules. It is also called gel permeation or gel filtration chromatography but the UIPAC recommended name is SEC. SEC is commonly used as a purification step and associated with Multiple Angle Light Scattering (MALLS), to estimate polymer molecular weight (see next chapter).

7.1 Principles and implementation of SEC

1. SEC column - RNase free equilibration The column for size exclusion chromatography is composed of a rather inert matrix of polymer particles that contain many pores of controlled size.

Implementation SEC was performed with an analytical Superdex S75 10/300 GL column (GE Healthcare). This column is composed of a spherical composite of cross-linked agarose and dextran with an average particle size of 13 μm. The column was pre-equilibrated with a phosphate buffer (23.5 mM potassium phosphate pH 6.5, 100 mM KCl, 10 mM $MgCl_2$ and 5 mM β-mercaptoethanol) prepared in RNase free conditions with RNase free components. These RNase free conditions enable to purify protein for RNA interaction studies. The volume of this column (CV) was 24 ml. SEC column was previously washed sequentially with 1.6 column volume of RNase free water, 0.5 M NaOH, 2 M NaCl and RNase free water. Between sets of sequential experiments, the column was stored in an aqueous solution with 20% ethanol. Before each experiment, the ethanol was removed with 1.6 CV RNase free water and the column was equilibrated with 1.6 CV gel filtration buffer described previously. A constant 0.5 ml/min buffer flow rate during the equilibration was controlled using the "BioLogic" system (Bio-Rad)

2. Injection of the sample on the column A small volume (<1/40 CV) of the mobile phase containing the macromolecule under study is loaded on the top of the column.

Implementation A 50 μl (< 0.084/40 CV) and 500 μl (< 0.84/40 CV) injection loops were used to inject the sample for SEC coupled to MALLS or for SEC only respectively. Before the start of an experiment, the injection loop used was systematically washed sequentially with RNaseZAP

(Sigma-Aldrich), RNase free water (>40 ml) and RNase free SEC buffer.

3. Elution The pumping rate is set to be constant and assumed to be so, which is controlled with the pressure monitoring at the column output. That's why the time variable to describe the evolution of the chromatography is usually converted as a volume variable to describe the peaks characteristics. During the elution of the mobile phase, the average residence time of the molecules and complexes within the pores will depends on their average shape and size. Small molecules will tend to be more effectively trapped by the pores than large molecules that will more likely migrate between them. Molecules larger than the pore size will be excluded from the pores and will elute first at an unresolved volume called the dead volume of the column or exclusion limit. At the other extreme, molecules smaller than the smaller pores will all elute the latest, at the same time. This is called the permeation limit. For molecules that elute between the two limits with different hydrodynamic properties, the SEC column implies a difference of the length of the average path followed by the molecules and so different elution volumes. Since the molecules were injected almost at the same time, species with similar hydrodynamic properties will elute almost at the same time. SEC does not concentrate the molecules during the elution.

Implementation The elution was performed at a constant 1 ml.min^{-1} flow rate. The value of the dead volume of the column was estimated using a solution of Blue Dextran ($\approx$ 8ml) of molar weight of about 2M Da.

4. Signal recording: UV, refractometry and MALLS During the elution of the mobile phase the liquid going out of the column can be analyzed using various type of detectors including UV, Multi-Angle Laser Light Scattering (MALLS), Viscometer, Refractometers and Mass Spectrometry.

Implementation For SEC only experiments, the chromatography was monitored with an UV detector working at 280 nm and 260 nm. Other detectors were used for the SEC-MALLS configuration (see next chapter).

7.2 SEC data analysis

Contributions to the elution volume of a species in SEC : The main variable of SEC-UV chromatography is the time dependent UV signal of the mobile phase. The peaks correspond to the elution volume of the UV absorbing species. Elution volumes at the exclusion or permeation limit indicate that the chromatographic column used is inappropriate to resolve the corresponding species. The SEC method is empirical with no well defined theoretical foundation that can quantitatively explain the elution volume in the fractionation range of a given species from a set of its properties. Two kinds of phenomena are known to contribute to the elution volume in the fractionation range of a given column :

- *Exclusion mechanisms.* This is the process already mentioned, the key process of SEC. It is difficult to link quantitatively the exclusion behavior of the molecules to a well defined biophysical parameter despite many articles have been published on this topic. The viscosity radius appeared as a good candidate compared to many others including mass, second virial coefficient, diffusion coefficient, etc (225). It is sometimes used in the so called universal calibration (159) and modified universal calibrations but further studies have found exceptions to them.

- *Potential adsorption* There are undesirable potential non size exclusion mechanism corresponding to the adsorption between of the molecule on the stationary phase. The stationary phase of the SEC used is made of cross-linked agarose and dextran. Some proteins are able to interact with the glucose residues.

Applications of SEC in this thesis As a result of the lack of clear theoretical foundations of SEC, a rigorous quantitative experimental analysis of the elution volume could not be performed. An experimental quantitative analysis would require using a calibration with other macromolecules. However such a calibration is of limited information since the choice of an appropriate set of proteins would require prior knowledge of the related properties for the protein under study. These calibrations have not been used in this work. However, SEC can be used quantitatively to estimate the protein purity. It can also be used qualitatively using a series of SEC profiles performed at different concentrations to assess whether a resolved species self-associate in the concentration range used. Indeed, if the populations of free and bound macromolecules are significantly different in the concentration range used, there will be a shift of the peak or multiple peaks with varying intensities. SEC was used in this work for these two qualitative and quantitative applications.

7.3 Non-trivial indiscernible processes in SEC

Five important and compatible non-trivial indiscernible processes in SEC are listed here :

- Both a single monomeric macromolecule, a highly stable complex or macromolecules involved in a fast association equilibrium compared to the duration of the SEC experiment can lead to a single peak in the chromatogram. Working at multiple concentrations can help to discriminate these situations
- Differences or similarities of elution might not be only related to exclusion processes (*cf* adsorption process above).
- Multiple conformers of a species with different exclusion properties in fast exchange compared to the duration of the SEC experiment won't be detected. The elution volume will depend on the hyrodynamic properties of the conformers and on their populations.
- Very large fibrillar aggregates may not be able to enter the column and thus may not be detected.

- If an equilibrium is going on in SEC, a quantitative study of resolved populations at equilibrium might not reflect the real thermodynamic behavior in solution because of the influence of the column matrix that segregate components during the elution.

Chapter 8

Principles, implementation and interpretation of SEC-MALLS

Definition and main biophysical applications of SEC-MALLS In static Multiangle Laser Light Scattering (MALLS) experiments, a monochromatic light scattered from a system is measured as a function of the angle between the incident beam by a set of detectors. This technique is also called classical scattering or total intensity scattering. MALLS measurements can give estimations of molar mass, molecular root mean square radius and information on intermolecular interactions

8.1 Principles and implementation of SEC-MALLS

1. Multi-angle light scattering detector - implementation Because of the variability of MALLS detectors design, I directly describe briefly here the system used, a DAWN EOS (Wyatt Instruments) with a laser with a vacuum wavelength of 690 nm and 18 detector positioned at different angles, measured from the oriented axis of the laser beam (**fig. 8.1**).

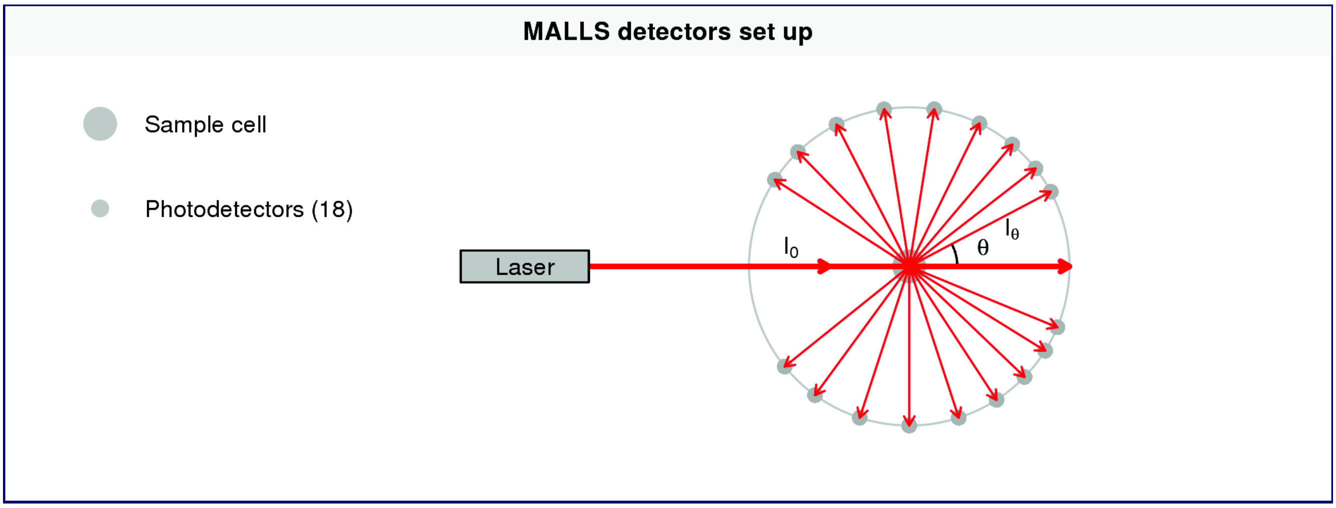


Figure 8.1: Photodetectors setup for the MALLS DAWN$^{®}$. EOS$^{\text{TM}}$ respecting real angles and with the notations of **tab. 8.1**(1).

This MALLS detector was used in line after a SEC chromatography step with the same parameters than the one described in the previous chapter. An in-line Differential Refractive index Detector (DRI) (RI 2000, Spares) was also used after the MALLS detector to follow the differential refractive index of the SEC eluate relative to the solvent (1.34), with the following set-up: [SEC at $\approx$ 25°C] $\longrightarrow$ [UV detector at $\approx$ 25°C] $\longrightarrow$ [MALLS at 25°C] $\longrightarrow$ [DRI]. The basic notations used in the MALLS part are given in **tab. 8.1**

Table 8.1: Notations used in the following MALLS equations

Notation	Definition
V	volume of the solution that scatter
N	number of particles in the volume V
N_a	Avogadro constant
c	mass concentration (kg/l) of the molecule in V
η	number of segments in the volume V (see **section 8.2**)
n_{mol}	refractive index of the solvated molecule
n_{sol}	refractive index of the solvant
λ_0	vacuum wavelenght of the incident wave
I_0	intensity of the laser beam
θ	angle between the incident beam and the detectors
d_θ	distance between the volume and the detectors
I_θ	scattered intensity at the angle θ
$d\Omega_\theta$	solid angle of the collected scattered light of the detector at angle θ
n_g	refractive index of the glass of the measurement cell
n_s	refractive index of the solvent used in the study
n_{ref}	refractive index of the reference compound used for the MALLS calibration
R_θ	excess Rayleigh ratio of the solute under study
a	characteristic radius of the molecule

2. Laser lamp and detector equilibration - implementation This step needs one hour for the DAWN system used (1).

3. Calibration and normalization of the MALLS detector The signal measured in a static light scattering experiment is mainly the angular dependency of the scattering intensity I_θ. For a solute, the signal can be quantified as the excess Rayleigh ratio of the solute under study. This ratio is defined here as the excess of scattering of the molecules under study over the scattering of the solvent. For a vertically polarized light it is defined with scattering intensities coming from a volume V into a detector with :

$$\forall\ \theta \in [1, 2\pi], \quad R_\theta = \frac{(I_\theta - I_{\theta\ solvent})\ d_\theta^2}{I_0\ V} \tag{8.1}$$

As the measured volume depends on each detector in a way difficult to precisely quantify, related to the geometry of the cells, the estimation of R_θ is done using an empirical relation between R_θ and I_θ :

$$\forall\ \theta \in [1, 2\pi], \quad R_\theta = f_{cal,\theta}(n_{ref}, n_s, n_g)\ \frac{(I_\theta - I_{\theta\ solvent})}{I_0} \tag{8.2}$$

This relation requires the use of the calibration of one detector at 90° with a chemical of known Rayleigh ratio and refractive index n_{ref}. It is followed by a normalization of all the detectors with an isotropic scatter in the solvent used for the study at the same pressure and flow rate (1). This procedure enables to estimate the calibration function $f_{cal,\theta}(n_{\text{ref}}, n_s, n_g)$. The exact procedure is device dependent. Since $f_{cal,\theta}(n_{\text{ref}}, n_s, n_g)$ does not depends of the scattering molecule, the calibration can be applied to any other system.

Implementation The calibration was done with HPLC grade toluene (filtered), a solvent with a high Rayleigh ratio. The calibration was checked with bovine serum albumin, a well characterized protein.

8.2 Data analysis of MALLS signal and related biases

Laser diffraction for MALLS can be conveniently studied non quantum mechanically using the Maxwell equations. An approximate analytical solution of these equations for MALLS has been developed and is widely applied to biological macromolecules. It is not the goal here to derive the MALLS equation from the Maxwell equations. However, I list in the following part the approximations that are used to derive the main equations used to fit SEC-MALLS data. These approximations mainly come from the seminal works of Zimm and Wyatt (307; 286) and are still widely used. However many approximations were criticized by their authors themselves and require discussion for a proper analysis of MALLS datasets.

System Let's consider within the measured volume slice a monomeric macromolecule or a stable oligomer characterized by a molar mass M. The following development holds true under this assumption of monodispersity. A simple adaptation of the results to the case of a mixture of multiple species of distinct molar masses is given afterwards. The incident light is polarized vertically. The N molecules are considered to be composed of a set of η scattering dipoles scatters called segments (286).

Hypothesis 1 The first condition required for the following development is that the size of the macromolecules studied is below λ_0 i.e. 690 nm. This limit corresponds to very large macromolecules and is easily checkable from the experiment results. This condition of validity is noted $\mathcal{A}_{MALLS}1$

Hypothesis 2 and consequences: Rayleigh-Gans-Debye approximation This approximation is: the incident wave is unaffected by the scattering molecules (286). With the notations of **tab.**

8.1, the Rayleigh-Gans-Debye approximation ($\mathcal{A}_{MALLS}2$) can be defined quantitatively :

$$\begin{cases} |\frac{n_{ref}}{n_s} - 1| \ll 1 & \text{molecular and solvent refractive indexes differences are negligible} \\ 2a\frac{2\pi n_s}{\lambda_0}|\frac{n_{ref}}{n_s} - 1| \ll 1 & \text{total phase shift of the incident wave through the sample is negligible} \end{cases} \tag{8.3}$$

Under the Rayleigh-Gans-Debye approximation, the excess Rayleigh ratio at an angle θ in a volume $d\tau$ with $\overrightarrow{s_\theta}$ being the difference between the unit vectors of incident and scattered rays, and $\vec{r}$ the vector between successive scattering segments in the volume $d\tau$ can be expressed as (286) :

$$R_\theta = \left(\frac{4\pi^2 n_0^2 \left(\frac{dn}{dc}\right)^2}{N_A (\frac{\lambda_0}{n_0})^4 c} \right) \left(\frac{\eta^2 N^2}{V^2} \right) \int \rho(r) \exp(2\pi i \overrightarrow{s} \; \frac{\overrightarrow{r}}{\frac{\lambda_0}{n_0}}) d\tau \tag{8.4}$$

The integration is over all orientations and magnitude of r at constant s.

To my knowledge, the application of the RGD approximation to biomolecules has not been studied generally. Although obviously useful, it seems that it introduce errors. A.R. Jones (132) mention three studies on spheres showing that if a signal error under 10% is expected with the RGD approximation, the following conditions are required :

$$\begin{cases} \frac{n_{mol}}{n_{sol}} \leqslant 1.25 \\ 2a \, \frac{2\pi n_0}{\lambda_0} \rightarrow 0 \end{cases} \tag{8.5}$$

Here the refractive index of the solvent is 1.34. The refractive index of the protein is unknown but is usually between 1.5 and 1.65 (115) for proteins. Assuming the highest value of 1.65, we have $\frac{n_{mol}}{n_{sol}} = 1.23$. This suggests that the accuracy may not be higher than 10% for some spherical proteins.

The second condition of the Rayleigh-Gans-Debye approximation means that polarization effects due to the anisotropy of molecules should not occur. This seems a more limiting assumption than the previous one, especially for the highly asymmetric RNA molecules. Anyhow it seems more relevant for small macromolecules, like the ones used in this work, than for very large ones (223).

Hypothesis 3 : molecules composed of independent scattering segments ($\mathcal{A}_{MALLS}3$). The relative position of two segments is characterized by the vector $\vec{r}$. The radial distribution function of the segments in a molecule corresponds to the probability of finding a segment at a distance r from another one in a volume $d\tau$. It is the sum of two disjoint contributions: two segments belong to the same molecule or they belong to distinct molecules. The probability P of finding one segment at a distance r from another one in a volume $d\tau$ is easily expressed as the sum of the probabilities of the two scenarios :

$$P = (\eta \rho_{inter}(r) d\tau) + (\eta \rho_{intra} d\tau) \tag{8.6}$$

The internal density can be expressed using the normalized intensity distribution function (307) :

$$P(\theta) = \int \rho(r) exp(2\pi i \ \vec{s}\frac{\vec{r}}{\lambda}) d\tau \tag{8.7}$$

Hypothesis 4 and consequences: Single point intermolecular contact restriction and concentrations infinitely small ($\mathcal{A}_{MALLS}4$). In the previous equations, the expression of the external density terms is harder. Under the hypothesis that two molecules can interact only at one single contact and with infinitely small concentrations, we have, with X a unexplicited term :

$$\int \lim_{c \to 0} [\rho_{inter}(r)] exp\left(\frac{2\pi i \vec{s}\vec{r}}{\lambda}\right) = \eta^2 X P^2(\theta) \tag{8.8}$$

This hypothesis is due to Zimm that recognized that this was the approximation most likely not valid in the MALLS equation derivation (307). As a consequence of this approximation, the SEC-MALLS cannot be used rigorously for the study of a mixture of particles with transient intermolecular large site interactions in the time scale of the experiment (1 data/seconds recorded). The potentially complexes and biased species separation from the SEC add also a complication. This is why the SEC-MALLS method was not applied in this work to study the protein/RNA stoichiometry and affinity. If the concentration is high, other terms should be added in this expression but there are difficult to describe quantitatively (286). In the eluate of SEC, the concentration of each slice is different from the others and is not "infinitely small". It is however widely assumed that these concentrations are sufficiently small for this approximation to be useful.

Summary: Debye-Zimm equation Under the hypotheses $\{\mathcal{A}_{MALLS}\ i, \ \ i \in \{1,2,3,4\}\}$, the equations previously given can be combined in the so called Debye-Zimm equation. The excess of scattering of the molecules under study over the scattering of the solvent at an angle θ in a volume $d\tau$ can be expressed like this (286) :

$$\begin{cases} R_\theta = \left(\dfrac{4\pi^2 n_0^2 \left(\frac{dn}{dc}\right)^2}{N_A(\frac{\lambda_0}{n_0})}\right) McP(\theta) \left[1 - 2\left(-\frac{-N_a \eta^2 X}{2M^2}\right) MP(\theta)c\right] \\ P(\theta) = \int \rho(r) exp(2\pi i \vec{s} \ \frac{\vec{r}}{\frac{\lambda_0}{n_0}}) d\tau \end{cases} \tag{8.9}$$

where X is a term not explicited here representing short range interactions. Importantly, this result has been derived with no assumption on the molecular shape/conformation of the molecule under study. The use of this equation requires to explicit the normalized intensity distribution function $P(\theta)$ also called the form factor (3). It can be expressed generally as a polynomial development :

$$P(\theta) = \sum_{j=0}^{+\infty} (-1)^j \, \alpha_j \left[2k \sin\left(\frac{\theta}{2}\right)^{2j} \right] \tag{8.10}$$

with the α_j being proportional to $\int r^{2j} \, dM$. Interestingly, the second coefficient α_1 has an easy interpretable value :

$$\alpha_1 = \frac{1}{3} \, r_g^2 \ \text{ with : } r_g^2 = \frac{1}{M} \int r^2 \, dM \tag{8.11}$$

where r is the distance from the center of gravity g of the molecule. r_g is called the mean square radius of the particle.

Practical use of the Debye-Zimm equation in this work The Debye-Zimm equation has two unknown parameters, the calibration constant already discussed and the increment of refractive index. The MALLS signal R_θ depends on the square of the differential refractive increment $\left(\frac{dn}{dc}\right)^2$. However, with the concentration of polymer measured with the Differential Refractive Index (DRI) detector, the concentration term c in the Debye-Zimm equation is replaced by $c = \frac{RI_{signal}}{\frac{dn}{dc} K_{RI}}$ which lead to a linear dependence of the MALLS signal on $\frac{dn}{dc}$ (272). The error on $\frac{dn}{dc}$ has thus much less effects with the use of a DRI detector. The estimation of $\frac{dn}{dc}$, required to analyze the MALLS data, can be performed by multiple ways :

- Direct experimental determination of $\frac{dn}{dc}$ usually require large quantities (a few mg) of macromolecules and dry weight measurements to be able to accurately determine the relation between the refractive index measured with a DRI and the macromolecular concentration. This direct method is thus often impractical for biological macromolecules for which such quantities are very difficult or costly to obtain.
- It could also possible to estimate $\frac{dn}{dc}$ using the knowledge of the concentration of the sample before the SEC-MALLS experiment and assuming a total recovery of the material going out the column but this is a strong assumption (possible adsorption in the column, imperfect injection, ...) so it can lead to errors that reach 20% (288).
- Consensus values. Many measurement in the literature have suggested that the distribution of $\frac{dn}{dc}$ was very narrow. An extensive theoretical study on human proteins has shown that the $\frac{dn}{dc}$ at 589 nm has an almost Gaussian distribution with a mean of 0.190 ml.g^{-1} and a standard deviation of 0.003 ml.g^{-1} (304). Thus a common approach is to use "consensus" values of about 0.185 in phosphate buffer saline at 25°C (278). According to a theoretical study on the $\frac{dn}{dc}$ values for human proteins (304), using the consensus value can lead to a maximum error of 10% for unmodified proteins.
- A study proposes a theoretical calculation of $\frac{dn}{dc}$ from the protein sequence (304). The prediction uses however strong assumptions in its derivation using in particular the Lorentz-Lorenz equations, valid for a few crystal structures, a point not discussed in the article. The proposed

approach applied to the MALLS conditions (25°C, laser wavelength of 690 nm and solvent refractive index of 1.34) lead to 0.1777 ml.g^{-1} for both TRBP-D1D2 and TRBP-D2 and 0.1776 ml.g^{-1} for TRBP-D1.

Thus using the calculated value will lead to a molecular weight estimations 4% lower than with the one estimated with the consensus value in phosphate buffer at 25°C. Since this error is moderate, the choice of one of the two approaches for TRBP constructs will not introduce major differences in the molecular weight estimation for monomeric macromolecules. To avoid the use of the strong assumptions of unknown effects in the theoretical calculation of $\frac{dn}{dc}$, the experimentally supported consensus values of 0.185 at 690 nm and 25°C for the protein constructs, leading to a potential molecular weight estimation error of 10 %, was preferred in this study.

A lack of information on $\frac{dn}{dc}$ for RNA does not enable to use very appropriate values. The same value than the one of the protein has been used in this study.

Adaptation for multiple masses Since the elution from a SEC peak can be composed of different species of distinct masses, the precedent equations can be expressed in this more general case by replacing the mass M by the a mass average $\overline{M}$ using data from each fraction i, here the "weight"-average (286) :

$$\overline{M_w} = \sum_i c_i M_i / \sum_i c_i \tag{8.12}$$

Fitting of the Debye-Zimm equation - implementation The data are then globally fitted for each SEC peak with the Debye-Zimm equation using the previous expression of $P(\theta)$.Different arrangement of the Zimm equation can be used to perform the fit of the data (Debye, Zimm and Berry models). Molecular masses were estimated here with the Debye model using ASTRA software version 5.3.4.20 (Wyatt Instruments). As $P(\theta)$ is an infinite series, the fitting is done with the first terms ($j \in \{1,2,3\}$) and the different models can be compared with model comparison techniques. The first two terms are usually sufficient for fitting molecules with r_g < 30 nm (it was the case in this study) but for rod-like molecules inclusion of the second order term in $\sin(\theta/2)$ can be required for the characterization of data collected at high angle even for much smaller molecules (286). The variables to be fitted are thus these $\alpha_{i,i\in\{1,2,3...\}}$, the weight average molecular mass $\overline{M_w}$ and if included, the second virial coefficient A_2. This coefficient was set to 0 in this study.

Thus one checkable hypothesis ($\mathcal{A}_{MALLS}1$) and 3 approximations ($\mathcal{A}_{MALLS}\ i,\ \ i \in \{2,3,4\}$) are required for the use of the MALLS equations to be fitted for the estimation of molar masses. These last three approximations are obvious sources of bias in MALLS. According to the previous discussion, it seems reasonable to estimate the maximum error on the estimated mass to range between 10 % (because of the assumptions in the Debye-Zimm equation derivation) and up to 20% because of the potential errors on the increment of refractive index, although the estimation might be more accurate for some systems. The imprecision on $\frac{dn}{dc}$, if it is not accurately measured experimentally, can be critical to discriminate between large multimeric species whose mass difference can be small. Independently to the previous potential biases, some errors can come from the noise of the data and the fitting process. They are reported here from the ASTRA software version 5.3.4.20 fits, taking into account the noise in each detector and the fitting procedure (3).

Other parameters like the root mean square radius is poorly determined form small macromolecules below 10 nm / 50 kDa. Their accurate measurement would require to have more signal so higher concentration or slower elution rates.

8.3 Non-trivial indiscernible processes in SEC-MALLS

Some important MALLS only non-trivial indiscernible processes are listed here :

- {A monomeric species} and {a mixture of species of different molar weight eluting at the same time} from SEC can lead to the same estimated average molar mass. Despite the SEC separation before MALLS is a key step to diminish the polydispersity of the samples measured, there is still a risk to have multiple species analyzed at the same time and thus a bias on the estimated mass.
- An imprecision on the increment of refractive index or a significant biais linked to the approximations 1,2 and 3 can impede the discrimination between various possibilities of larges complexes with similar molar masses.

Chapter 9

Principles, implementation and interpretation of SDS-PAGE

Definition and main biophysical applications of SDS-PAGE SDS-PAGE stands for Sodium Dodecyl Sulfate (SDS) PolyAcrylamide Gel Electrophoresis (PAGE). Gel electrophoresis enables in parallel high resolution separation of many macromolecules in a mixture. SDS-PAGE method is highly used for characterizing macromolecular sample homogeneity. The method used here correspond to a slightly modified version of the one described by Laemmli (153).

9.1 Principles and implementation of SDS-PAGE

1. Sample preparation for SDS-PAGE. We consider in this paragraph either proteins or nucleic acid that will be called generically macromolecules. The samples SDS-PAGE are prepared by heating macromolecules in a Sodium Dodecyl-Sulfate (SDS) solution with a high concentration of reducing agent. SDS (CH_3-$(CH_2)_{10}$-CH_2OSO_3-Na^+) is an anionic detergent that triggers a temperature catalyzed denaturation of the macromolecules. The reducing agent reduces the disulfite bounds.

Implementation A 4× solution for the sample preparation contained 128 mM Tris at pH 6.8, 10% glycerol, 2% SDS, 730 mM β-mercaptothanol and a small amount of bromophenol blue and was used at a 1× final concentration. The samples were heated at 95°C during 5 minutes. To remove potential aggregates, heated sample were then centrifuged during 10 minutes at 25°C and 18000 g just before loading them on the gel.

2. Gel preparation for SDS-PAGE Polyacrylamide gels are formed by polarization of cross linked acrylamide and bisacrylamide. The gel has globally uniform pores, whose size is dependent on the percentage of acrylamide. Polymerization with a comb enables to make wells in the stacking gel.

Implementation Gels were prepared in this work using two 1 mm staked gels 8 × 7 cm gels with 15 wells, a staking gel with 7.5% of $\frac{29}{1}$ acrylamide/bisacrylamide for sample loading and concentration, and a running gel with 15% of $\frac{29}{1}$ acrylamide/bisacrylamide for the macromolecules separation. The composition of the staking and running gel solutions for two gels are given in **tab.**

9.1 for a two gel preparation. The TEMED (Tetramethylethylenediamine) and APS (Ammonium persulfate) are added and mixed extemporaneously just before the gel casting in a Mighty small dual gel caster device. The staking and running were poured successively in the caster device with an time interval long enough for the running gel to polymerise. A solution of isopropanol was added during this interval at the top of the running gel to avoid surface irregularities. The alcohol was then removed with an abundant water wash. The electrophoresis was performed at 180 V using a SE 250 Mighty Small II system.

Table 9.1: Composition of the running and staking gel solution to prepare 2 80 × 70 × 1 mm gels.

Solution	Vol. running gel (ml)	Vol. staking gel (ml)
water	4.125	4.4
Tris 2M pH 8.8	2.05	0
Tris 1M pH 6.8	0	0.73
SDS 10% (w/v)	110.10^{-3}	55.10^{-3}
Glycerol 50% (v/v)	143.10^{-3}	0
Acrylamide/bisacrylamide $\frac{29}{1}$ 40%	4.125	826.10^{-3}
TEMED	11.10^{-3}	$10.3.10^{-3}$
APS 10% (w/v)	80.10^{-3}	30.10^{-3}

3. Electrophoresis Samples are loaded in the stacking gel wells. An electric field is then applied to the gel with the positive charge being at the opposite position of the samples to trigger their migration. The protein-SDS complex depends mainly of the length of the peptidic chain (about 1.4 g SDS/g protein i.e. about 1 SDS molecule per 2 residue (221)). The macromolecules-SDS complexes being highly negatively charged, it is considered that the main factor that control the migration is the length of the polymer, which is the property of interest when performing SDS-PAGE. The chemical composition has also an incidence. It has been shown that acidic and basic proteins migrate not as expected taking only into account their chain length. Acidic proteins (negatively charged) tend to bind less SDS (negatively charged) than the average, which results in a higher mobility. It is the contrary for basic proteins (e.g. Histone H1, troponin 1) (110). Additionally, proteins with high proline content have a decreased mobility probably as a result of increased rigidity and kinks (110). As a result of these phenomena, an interpretation of the value of the SDS-PAGE protein mobility by its molar mass in order to estimate it gives errors of a least 10 % but they can reach 30 % (110). Moreover, non full denaturation state of the macromolecule during the migration is also expected. It has been shown by NMR that polypeptidic chains, even in the presence of strong denaturants like 8 M urea, guanidine hydrochloride or SDS are not completely random coils. So the chemical composition is also likely to impact these conformation and the mobility.

Implementation The electrophoresis was performed at a constant voltage of 180 to 220 V in a migration buffer (25 mM Tris, pH 8.3, SDS 2 g/l and glycine 14.4 g/l).

4. Macromolecule detection - electrophoretic mobilities After the end of the migration, the macromolecules are commonly detected by staining. A solution of Coomassie blue is often used for its convenience. The stainability of a protein with Coomassie blue band depends on its chemical composition (110), on the density of macromolecules in the band (206) and the staining increases with time. As a result of the time dependence property, for quantification purpose it is required that the set of macromolecules used for calibration are stained simultaneously as the other lanes to be studied.

Implementation The gels were incubated with a fixing solution (10% (v/v) acetic acid, 45% ethanol and 45% water) during 10 minutes, stained (10% (v/v) acetic acid, 45% ethanol, 45% water, 600 mg/l Coomassie Blue R250) and destained with a solution of 10% acetic acid and then stored in distilled water. The different species detected by the staining solution can be seen as bands that can be characterized by their electrophoretic mobility μ. With E the electric field (V.cm^{-1}) and v the migration rate (cm.s^{-1}), it is defined as: $\mu = \frac{v}{E}$. The mobility of a sample of interest is often compared to the ones of known macromolecules mixtures called standards.

9.2 Non-trivial indiscernible processes in SDS-PAGE

Some of the main non-trivial indiscernible processes in SDS-PAGE are listed here :

- {A pure protein sample} and {the same sample with additional very large aggregates} that can't enter the gels will lead to the same apparent result. The systematic addition of the coloration of the stacking gel can help to detect such aggregates. Lack of detection can also happen with small contaminants like small peptides that can be eluted from the migration or lost during the fixing/staining step. Thus, the lanes might not reflect the full macromolecular composition of the sample.

- {A basic protein} and {a smaller neutral protein} can lead to two bands with the same electrophoretic mobility. Many similar scenarios based on the multi-factorial determinants of the electrophoretic mobility can be imagined.

- {A pure sample with two different migrating species in the gel linked to imperfect denaturation} and {a sample with two distinct species in solution} can lead to the same indistinguishable two band result.

- {A monomeric protein}, {a homo-oligomeric protein} or {a protein covalently linked to other undetected macromolecules} can all elute with one band. A one band signal does not give information on the covalent or non covalent macromolecular assembly state in solution.

Chapter 10

Full cell ITC for the study of chemical equilibrium

Definition and main biophysical applications of ITC UIPAC proposes the following definition of Isothermal Titration Calorimetry (ITC) (244): "an ITC measurement consists of monitoring the transfer of heat between an analyte solution in a sample vessel and a reference solution in a reference vessel upon injection of a small aliquot of titrant solution into the sample vessel at a fixed ITC operating temperature". Compared to many biophysical methods, calorimetry is an old field that started with biological studies performed by Lavoisier and Laplace in the 1780s on guinea pigs (158). Calorimetry has received in increasing interest in molecular biophysics linked to the development of increasingly sensitive calorimeter with small sample requirements. ITC is widely used to determine thermodynamics parameters of biological interactions between combinations of proteins, nucleic acids and small molecules.

10.1 Principles and implementation of full cell ITC

I describe here a typical full cell calorimeter and how is generated the signal from a mixture of two solutions. The full cell titration starts from a cell filled at the beginning of the experiment. The injection of liquid in the cell during the titration leads to the ejection of a similar liquid volume out of the cell.

1. Calorimeter Let's consider the monitoring of the mixing of two solutions using a full cell ITC. I describe here the main parts **fig. 10.1**. The calorimeter contains two lollipop shaped cells, a sample cell and a reference cell. Cells are made with an inert material and have heaters on their flat part. Their inner volume range between 2 ml and 200 μl depending on the calorimeters; the inner volume of the syringe is much smaller. The two cells are placed in an adiabatic shield regulated at the desired temperature T_{exp}. Both cells are connected to the outside of the adiabatic shield by long and thin access tubes. The mains steps in the ITC set-up to the signal recording are described in the following part.

Implementation The full cell ITC system used in this thesis is an ITC-200 from MicroCal **fig. 10.1**. The inner cell volume is 201.4 μl while the syringe volume is 35 μl. Cells are in Hastelloy-

C. It is at the time of writing the most sensitive ITC apparatus available and the one using the smallest amount of liquid.

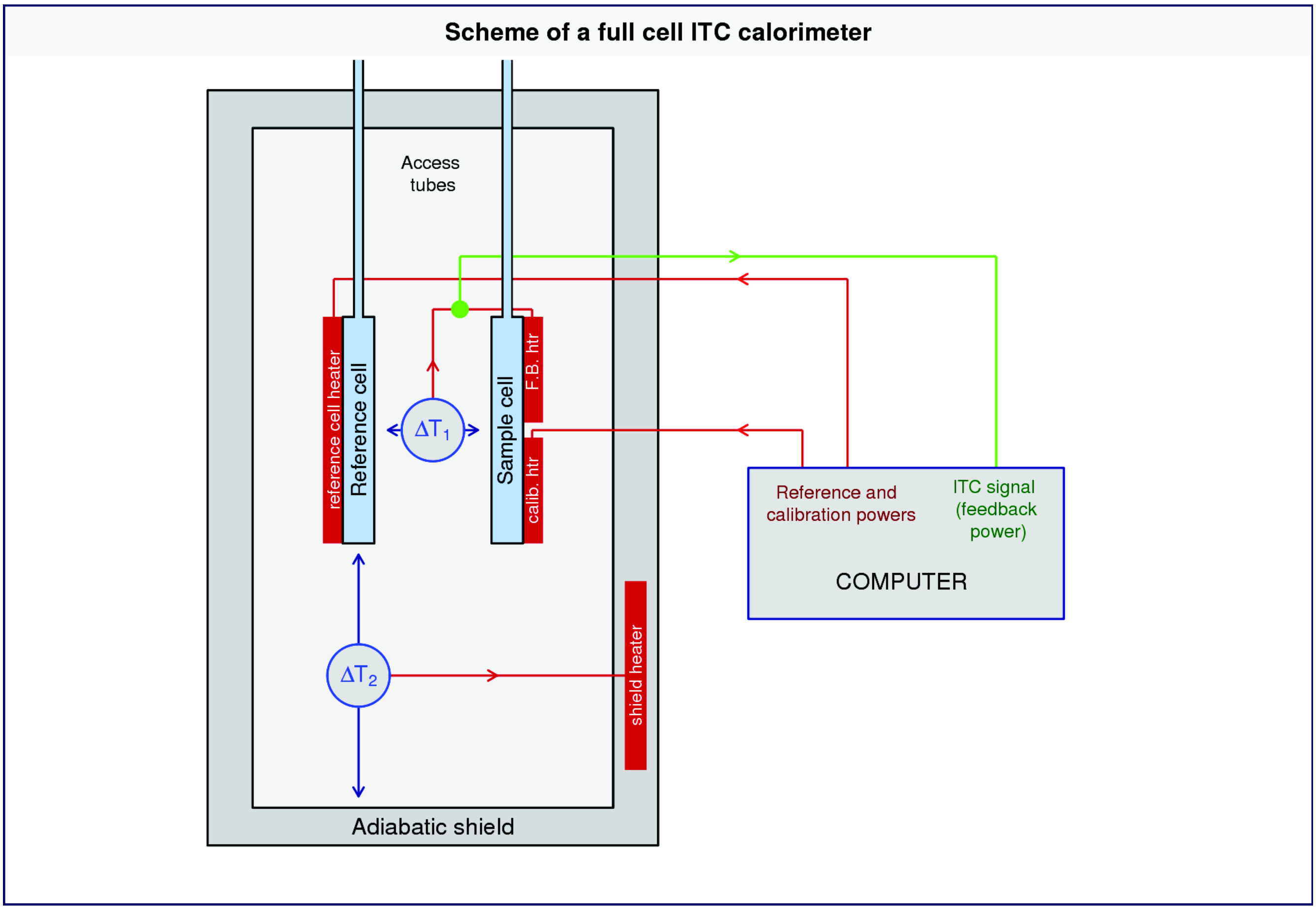


Figure 10.1: Simplified scheme of a full cell calorimeter realized from the original description of Wiseman *et al.* (280). This scheme corresponds to the calorimeter used in this study. The heaters (htr) are colored in red. Thermopiles that measure temperature difference are represented in blue circle with the name of the parameter measured in the circle and arrows pointing to the calorimeter elements compared. The injection and homogenizing syringe in the sample cell is not represented. Cal.: calibration. F.B.: Feed-Back

2. Wash The sample cell and the injection syringe are part of the calorimeter so in order to avoid adding reactions to the potential ones that are the goal of the study, these elements require to be extensively cleaned before the start of the experiment. This is particularly important for work with RNA since RNases are commonly used to calibrate the calorimeter (78).

Implementation In the present study, before a series of experiment, the sample cell and the injection syringe were washed using an incubation during 5 minutes in RNase ZAP (Sigma-Aldrich) diluted in RNase free water. An extensive wash was done using a long needled RNase free syringe with RNase free water, stopping a few washes after no longer moss due to the RNase ZAP (Sigma-Aldrich) was detected visually. This corresponds to 15 to 20 washes. Additional 5 washes were done with the RNase free buffer used to solubilize the macromolecules. The cell was stored filled by the buffer until the loading of the sample and the syringe was dried. To avoid inter-experiment contamination, at least 6 wash steps with both water and buffer were performed between each experiment.

3. Filling - implementation A long RNase free needled syringe is used to fill the cell and the access tube with the analyze solution. The reference cell is usually already filled indifferently with water or a buffer. The hypothesis that both cells are filled is required for the analysis so it is very important that there are no bubbles in the cell. That's why it is first necessary to degas the sample solution or to centrifuge it to remove bubbles in the solution. Moreover, it is required to add slightly more volume than the sum of the cell volume and the access tube volume to be sure that the system has been completely filled. The access tubes enable potential bubbles located at the top of the cell to go up. The syringe is filled as any syringe. The samples were only centrifuged here (18 000 *g* during 10 min at 25°C) since this avoids the risk to increase the concentration of the macromolecules by evaporating some solvent components.

4. Potential small non collected injection Before the insertion of the syringe into the cell, it might be useful to do a small injection to have a homogeneous movement of the syringe piston (199) to avoid special behavior of the first recorded point. This step was not performed here.

5. Insertion of the syringe into the cell and stirring After the filling, the syringe is placed into the cell. This action cause part of the solution in the cell to go out of the access tube so the volume of solution in the cell is now $V_{\text{cell}} - V_{\text{syringe in cell}}$. The syringe has a small paddle at his tip **fig. 10.1** that is used to homogenize the cell solution using a circular motor that rotate the whole syringe from this step to the end of the experiment. Stirring has to be sufficient for a fast mixing.

Implementation A 800 rpm stirring was chosen in this study.

6. Thermal equilibration of the cell The thermal regulation circuit works permanently but it will have a key role from this step till the end of the experiment.

- *Temperature control mechanism.* The temperature is controlled in the range [2-80°C] via a feedback mechanism. The adiabatic shield is regulated to the temperature T_{exp}. A thermopile connects the adiabatic shield with the reference cell and produces a signal proportional to their temperature difference ΔT_2. This signal is amplified and fed to the feedback heater on the jacket to maintain its temperature very close to T_{exp}. In parallel, to prevent heat leaks from the sample cell to the surrounding, a thermopile measure the temperature difference ΔT_1 between the sample cell and the reference cell, which is compensated by the sample cell feed back heater. This is the important signal monitored over time. Because the temperature regulation is performed by heaters, this regulation will work better when T_{exp} is higher than the room temperature; for lower temperature, a coolant circuit is used in the system (water or dry air).

- *Signal recorded.* To avoid recording noise, a constant power below one milliwatt is dissipated in the heater of the reference cell. This will trigger the cell feedback circuit described previously that will drive ΔT_1, back to 0 K. The signal of the "injected power" is recorded over time via a computer. In the absence of a temperature change in the cell, the recorded feedback power will be constant.
- *Calibration.* Accurate measurement of the power related to the solution mixing require a calibration correction by measuring the heats of standard chemical reaction and performing internal electrical pulses heating checks. The electrical correction is implemented via a constant power applied to the cell heaters. The correction is performed by a devoted heater on the sample cell. Ideally a standard reaction is used to check that the calorimeter is well calibrated.

It has to be underlined that contrary to the adiabatic shield, the temperature of the syringe is not controlled in the ITC-200 calorimeter and is at room temperature. So despite the titrant in the syringe arrow tip will have time to equilibrate at the cell temperature, there will be a thermal transfer between the room and the cell. This effect can be significant when the temperature of the room is highly different from the one of the cells (e.g. 61°C in (308)). So it is important to maintain the temperature of the room constant. Ideally about 30 minutes are required for the equilibration. This is assessed by waiting to the time required to have an almost constant baseline before the beginning of the experiment. For this purpose, before the start of the injections, the temperature difference between the cell and the jacket is monitored in live.

Implementation. The room temperature between 22°C and 23°C and at least 20 minutes equilibration were used to reach an almost flat baseline.

6. Isothermal titration and signal recording A motor drives the plunger to inject a controlled small volume amount from the syringe to the cell. The mixing is finished within a "few seconds" after injection, according to MicroCal "visual inspection" at 400 rpm (280). The observable in ITC is the input of power over time delivered to the sample cell feed back heater required to maintain its temperature equal to the one of the reference cell. If the mixing cause a change of the temperatures between the two cells, there will be a change of the feedback power followed by a progressive return back to the baseline. If the solution mixing is exothermic, it will create a negative contribution to the power signal measured; if the solution mixing is endothermic, it will create a positive contribution to the signal. If an exothermic or endothermic chemical reaction is happening, the shape of the peaks will have a fast almost linear part and a slower decreasing part. The delay between injections should be long enough for the baseline to go back to a constant value. Otherwise, the signal is not interpretable quantitatively. In ITC datasets of solution mixing associated with one or more reactions, the first injections will give large heats because the majority of the injected titrant, in small proportion, will bind to the analyte in excess in the cell. During the titration, binding sites of the ligand or the analyte will get saturated in the cell and the peaks area will decrease, not obviously to zero. At one point there will no longer be changes in the signal, indicating a saturation in the titration. The experiments have to reach this point in order to exploitable. A classical example of recorded signal, for an exothermic reaction, is given in the result part in **fig. 17.1**.

10.2 ITC data analysis for thermodynamic reaction characterization, related biases and proposition of a new systematic strategy

Two main motivations have led to the redaction of an "experiment design and data analysis" part for ITC, in a more comprehensive way than what has been done for the other techniques :

- *Lack of published demonstration of ITC fitting equations.* Despite ITC is widely used and that the principles rely on simple thermodynamics, to my knowledge, there is no published full demonstration of the widely reported ITC equations used to fit the experimental data. This might not appears as a big deal since they are not long to demonstrate but it appears that when reported into the articles, the ITC equations are stated without the explicitation of the hypothesis, approximations and boundaries that are necessary to derive these equations from basic thermodynamics. This is particularly striking for the widespread "single sites models", a special microscopic model that is often erroneously considered as a generally applicable model (actually restricted to one set of independent sites). Thus there is a risk to over/mis-interpret some ITC datasets. The full equations and their validity conditions are summarized in this chapter.
- *Explicitation of a reliable strategy to fit ITC data.* The most used ITC equation to fit the data is based on the one set of independent sites microscopic models. It provides affinity constants, reaction enthalpies and molar ratios. Additionally to the point raised in the previous paragraph, it is well known that the floating molar ratio parameters can be biased with error on the active concentration of the molecules under study. I emphasize here the approach used in the thesis using global fitting of the ITC data by successive series of multiple models compared with an information theoretic approach. This aims at determining the relevant information content of the ITC data and is implemented in the GNU R language (228).

10.2.1 Baseline correction of ITC datasets

Slight differences of temperature of the room during the experiment or underfilled cell, might cause continuous drift of the ITC signal baseline. So the first step in ITC data analysis is a baseline correction. Some manufacturer software encourages to manually adjust the baseline. However this practice can introduce a non reproducible user dependent bias. It can also encourage bad practice consisting in back-adjusting the baseline to improve the fitting for the selected model. Moreover manual baseline correction is not compatible with the high throughput processing of multiple datasets. Before 2012 no automated ITC baseline correction algorithms have been proposed in the literature. In mid-2012, a good baseline correction has been proposed, base on singular value decomposition (140).

Implementation. In the work presented here, the moderate baseline correction required to adjust the datasets were performed using a combination of two general baseline correction algorithms working well for the structure of the available datasets. R baseline package was used to implement these corrections (169). The first one is Asymmetric Least Squares Smoothing (ALS) (66). Using the terminology used in the latter article, this algorithm was used with $\lambda = 0$ and $p = 0.01$. This first correction worked well but was imperfect for the smallest injection at the end of the titration. A second correction was used to improve the first one, based on Iterative Restricted Least Squares (IRLS), an algorithm implemented in the R baseline package (169). Using the terminology used in the latter reference, this algorithm was used with $\lambda_1 = 0$, $\lambda_2 = 5$ and $w_i = -0.2$. It has to be underlined here that the choice of the parameters is not appropriate for any datasets, and has to be adjusted for different datasets. Thus, the proposed approach can be qualified as semi-automated baseline correction.

10.2.2 Secondary measurable from baseline corrected ITC data

After baseline correction, there are two types of data that can be extracted from ITC curves :

- *peaks integrals*. The integration of the power over time for each peak gives a vector of heat transfers with one component for each injection.
- *peaks shapes*. Kinetic studies of enzymatic reaction by ITC were pioneered by Julian Sturtevant (reviewed in (18)). Recently a group (25) has proposed to extract kinetics information from the peaks shapes in more general cases but the method is still at a development stage.

This part is devoted to the extraction of thermodynamic parameters from the vector of heat transfers. Let I be the number of injections and t_i the time at which the injection i end, t_0 is defined as the beginning of the titration. Let Q be the vector of heat transfers during the titration. For a given reaction involving the titrant, the magnitude of the heat transfers at each titration step will depend on the amount of titrant injected into the cell. The standardized vector of heat transfers q is obtained by dividing each component of Q by the corresponding number of moles injectant :

$$\left\{ q \in \mathbb{R}^I / \forall\, i \in [\![1, I]\!],\ q_i = \frac{Q_i}{n_{injectant}(i)},\ Q_i = \int_{t_{i-1}}^{t_i} P(t) \right\} \tag{10.1}$$

This standardization enables to compare easily the integrated data of two different experiments using different concentration of titrant. Modelization of the continuous ITC power signal (dq/dt) with ordinary differential equation can give interesting insight into the properties of binding models (280; 224). This approach is not used here where the more common modelization of the vector of heat transfer is used. The next part presents the thermodynamical modeling of the vector q.

10.2.3 Total concentration estimation at each titration step

Before starting the thermodynamic modelization of the system, it is necessary to calculate the total concentration of each titrant and analyte solution at each step of the titration. There is a least 3 different formulas found in the literature with no justification for most of them. They derivation depends whether it is considered that the expelled volume after each titration has a composition similar to the one present in the cell before the injection or whether the mixing is fast enough so that the expelled volume will progressively be charged by the injected material. A demonstration of the different formulas under their different hypothesis of validity is given in annex **section 20.4**

Implementation. In the experiments performed here, the injection last about 0.5 - 1 seconds and the spinning is performed at 800 rpm so about 13 rotations by second occur. It seems likely that in these conditions there will be a continuous mixing during the injection. The dilution will depend on the usually poorly known viscosity of the solutions and the geometry of the cell and stirrer. However, a full hydrodynamic study of the mixing is not required to solve this since in the range of injected volumes used, the two extreme case, "no mixing" and "mixing" will lead to very similar total concentration estimation in the common ITC conditions. Since the continuous mixing approach seemed more appropriate this is the one I choose to use in this thesis ($\mathcal{A}_{ITC}1$). The total concentration expressions are given in **section 20.4** and in **section 10.2.5.1**.

10.2.4 Energetic modelization of ITC heats transfers

Application of the first principle of thermodynamics Only the beginning of the demonstration is given here to underline some key approximations/hypothesis. The remaining part is available in the annex **section 20.4** We consider a reaction happening without gas production ($\mathcal{A}_{ITC}2$). At each injection of a volume v in the sample cell, a similar volume will be expelled in the collect bath. The resulting elevation of liquid in the collect bath is very moderate because its diameter is much larger than the diameter of the narrow access tubes **fig. 10.1**. The resulting increase of pressure in the calorimeter cell can thus be neglected. The pressure P is assumed to be constant from now ($\mathcal{A}_{ITC}3$).

According to the first principle of thermodynamic applied the volume corresponding to the inner part of the calorimeter cell, the variation of internal energy U, of kinetic energy E_k and the potential energy E_p are linked to the work W and to the thermal transfers Q exchanged with the outside with :

$$\Delta U + \Delta E_c + \Delta E_p = W + Q \tag{10.2}$$

The content of the cell does not undergo variations in kinetic or potential energies so: $\Delta U = W + Q$. Assuming no gas is released during the titration ($\mathcal{A}_{ITC}4$), V, P and T are constant. Since

$H = U + PV$ we have for the system: $\Delta H = \Delta U$ (107). So the measured thermal transfer is :

$$Q = \Delta H_{mixing} - W_{\text{mixing}} \tag{10.3}$$

The goal is then to express Q as a function of the thermodynamic parameters of the reaction, which require explicitation of the enthalpy and work terms :

Work contributions on Q The work of mixing corresponds to the work required to perform the titration. It has been shown (107) that the injection work was detected by ITC, leading to small positive contributions to the signal. This work can be interpreted as the energy required to compensate the energy dissipated by the friction of the fluids during the injection (107). But a contribution to the friction forces linked to the movement of the liquid in the long narrow access tube can also be conceived.

Enthalpy contributions on Q They can be classified into two groups :
- *Enthalpies linked to the potential reactions under study*. This is the enthalpy we usually want to measure. This enthalpy depends on the sum of surface weak forces, solvation changes and conformational changes involved in the reactions.
- *Enthalpies not directly linked to the reaction under study*: enthalpies associated to the dilution of the titrant and/or analyte if they are involved in an equilibrium with themselves, enthalpies due to the imperfect mismatch of the syringe and cell buffer and reaction with a buffer component (acido-basic reactions). In the latter case, the potential reaction under study occurs with the release or undertake of protons that will lead to parallel acido/basic reactions with the buffer system. The magnitude of these enthalpy variation contributions depends on the ionization enthalpy ΔH_{ion} of the buffer component(s) used and the number n of proton released: $\Delta H_{protonation} = n.\Delta H_{ion}$. The buffer ionization enthalpy can be high for some buffers like amine buffers (ΔH_{ion}[Tris, pH = 8.1] = 47.44 kJ.mol^{-1}) and much lower with others like phosphate (ΔH_{ion}[Phosphoric acid, pH = 7.2] = 3.3 kJ.mol^{-1}) (41).

As a result, the thermal transfers estimated from the ITC signal integration is *a priori* not directly linked to the reaction enthalpy of interest $\Delta H_{\text{target reactions}}$:

$$Q = \sum_{r\, \in\, \text{reactions}} \Delta H_r - W_{\text{mixing}} \tag{10.4}$$

with multiple reactions: *target reaction(s)*, buffer reactions, dilution of the analyte, dilution of the titrant, reactions linked to buffer mismatches and potential other reactions.

Isolation of the enthalpy of reaction Because of the previous relation, the estimation of an accurate reaction enthalpy require to isolate the reaction enthalpy terms by performing other experiments and/or ensuring that they are negligible. Various experiments of result to be subtracted or added to the vector Q of the main experiment are commonly described (41; 87). For the work term, a blank experiment, adding the titrant into the buffer and buffer into the titrant can be used to estimate this work W (41) or check that it is negligible compared to the measured enthalpies. For a study of a two molecules interactions, the following experiments are required to estimate the enthalpy terms (dil.: dilution ; buf.: buffer) :

- [Syringe: titrant | cell: buffer only]: $Q_{\text{dil. titrant}} = \Delta H_{\text{estim. dil. titrant}} + W_{\text{mixing dilution titrant}}$
- [Syringe: buffer only | Cell: analyte]: $Q_{\text{estim dil. analyte}} = \Delta H_{\text{estim. dil. analyte}} + W_{\text{mixing dil. analyte}}$
- [Syringe: buffer only | Cell: buffer only]: $Q_{\text{estim buf./buf.}} = \Delta H_{\text{estim. buf./buf.}} + W_{\text{buf./buf.}}$

It has to be underlined that the signal in these complementary experiments do not perfectly estimate the enthalpy terms, not only because of the presence of the work terms but also because of the fact the reaction are not exactly happening in the same conditions than in the main experiment. Indeed the absence of at least one component can change the properties of the reaction mixture like the viscosity which can have an effect on the dilution experiments for instance. This comment is also valid for the work term. If we neglect these effects, one can use the following expression to estimate the enthalpy variation of the target reaction or show that the correcting sum is negligible and would just add noise.

$$\Delta H_{\text{target reactions}} = Q_{exp} + Q_{\text{dilution titrant}} + Q_{\text{dilution analyte}} - 3\Delta H_{\text{estim. buffer/buffer}}$$

10.2.5 ITC macroscopic models to fit ML$_n$ complexes

Various models are always possible to fit to a curve or set of curves. In ITC, two kinds of thermodynamic models can be fitted to the data: the macroscopic models and the microscopic (i.e. sites specific) models. The first type represents the dissociation constants related to the sequential binding process whereas the second type of model corresponds to the site specific dissociation constants directly related to the kinetic constants.

$\mathcal{A}_{ITC}5$: I consider here a reaction involving complexes of type ML$_{m,\ m\in\mathbb{N}^*}$. Let n be the stoichiometry associated to the largest complex. I also consider that the reacting species are only composed of M and L.

Using the precedent equations and the expression of the concentrations, the classical thermodynamic calculations leads to the following equations with the notations in **tab. 10.1** :

10.2.5.1 Macroscopic modelization of reaction enthalpies for ITC experiments with ML$_n$ complexes

Table 10.1: Notations used in models of enthalpies of reaction

Notation	Definition
n	maximum stoichiometry of complexes ML_n
E	experiment $\in \mathbb{N}^*$
I_E	number of injections in experiment E
N_E	number of injection points used in experiment E ($N_E \leq I_E$)
i_E	injection of experiment E, $i_E \in [\![1, N_E]\!]$
v_{i_E}	injection volume in the injection i_E
$N_{E,i}$	number of moles of injectant in the experiment E injection i
V	effective volume of the cell ($V = V_{\text{cell}} - V_{\text{plunger}}$)
$[\text{X}]_{E,i}$	concentration at equilibrium of species X in experiment E at injection i ; $\text{X} \in \{\text{L}, \text{M}, \text{ML}_{m,\ m\in [\![1,n]\!]}\}$
$c(X)_{E,i}$	total concentration of species X in experiment E at injection i
M_0	initial concentration of species M
L_0	initial concentration of species L
$\overrightarrow{\Delta_r H_\lambda}$	vector of reaction enthalpy of binding step
$\Delta_r H_\lambda$	component of $\overrightarrow{\Delta_r H_\lambda}$ for step λ
$\overrightarrow{\beta_E}$	vector of constant baseline correction of length E

The formula required to fit the macroscopic sequential binding model under the assumptions $\mathcal{A}_{ITC}i,\ i \in [\![1,5]\!]$, called from now "ITC formula to fit macroscopic models", is :

$$\forall\ (k, i, E) \text{ in } [\![1, n]\!] \times [\![1, I]\!] \times [\![1, E]\!], \tag{10.5}$$

$$q(i) = \frac{V_e}{N_{E,i}} \sum_{j=1}^{n} \left(\sum_{\lambda=1}^{j} \Delta_r H_j \prod_{\lambda=1}^{j} K_\lambda \right) \left[\left(1 + \frac{v_{E,i}}{2V}\right) [\text{M}]_i [\text{L}]_i^j - \left(1 - \frac{v_{E,i}}{2V}\right) [\text{M}]_i [\text{L}]_i^{j-1} \right] \tag{10.6}$$

with :

$$c(L)_i = \begin{cases} \mathrm{L}_0\left(1 - e^{-\frac{v_{E,i}}{V}}\right), & \text{if L is injectant} \\ \mathrm{L}_0\left(e^{-\frac{v_{E,i}}{V}}\right), & \text{if L is in cell} \end{cases} \tag{10.7}$$

$$c(M)_i = \begin{cases} \mathrm{M}_0\left(1 - e^{-\frac{v_{E,i}}{V}}\right), & \text{if M is injectant} \\ \mathrm{M}_0\left(e^{-\frac{v_{E,i}}{V}}\right), & \text{if M is in cell} \end{cases} \tag{10.8}$$

$$[M]_i = \frac{c(M)_i [L]_i^j}{1 + \sum\limits_{k=1}^{n} \left(\prod\limits_{j=1}^{k} K_j \right) [L]_i^k} \tag{10.9}$$

$$[L]_i = \left\{ x \;\; /c(M)_i = \left[\frac{\sum\limits_{k=1}^{n} \prod\limits_{j=1}^{k} K_j x^k}{1 + \sum\limits_{k=1}^{n} \left(\prod\limits_{j=1}^{k} K_j \right) x^k} \right], x \in [0, c(L)_i] \right\} \tag{10.10}$$

The goal in the ITC data fitting with macroscopic models of type ML_n can then be summarized like this :

$$\underset{\overrightarrow{K}, \overrightarrow{\Delta_r H}, \overrightarrow{\beta}}{argmin} \left[\sum_{j=1}^{e} \left(\frac{1}{N_e} \sum_{\lambda=1}^{I_e} \left(\frac{q_{expe,E,i}}{N_{E,i}} - q_{theo,E,i} \right)^2 \right) \right] \tag{10.11}$$

10.2.5.2 Microscopic modelization of reaction enthalpies for ITC experiments with ML_n complexes

The richest models are the microscopic models since they enable to compare the behavior of each site of the macromolecule and to deduce easily the macroscopic constants. However, the possibility of relevant fitting site-wise models is restricted with the monodimensionnal ITC data, compared to NMR for instance. A molecule with different binding sites has different site-wise mechanisms, depending on 2 variables: *variable A*, identicity of the behavior of the sites (identical or different behavior) and *variable B*, independency of the sites (independent, positively cooperative, negatively cooperative). For a molecule with n binding sites, the number of these mechanisms $N_{mechanism}(n)$ is not usually reported. For n sites, there are n among n possibilities of choosing whether the n sites are identical or not (*variable B*), and for each possible pair of sites $\binom{n}{2}$ there are three possibilities of relation between the sites (*variable A*). The two variables A and B are independent so the number of possible mechanism $N_{mechanism}$ is the product of these values :

$$N_{mechanism}(n) = \binom{n}{n}\ 3\binom{n}{2} = \frac{3}{2}n^2(n-1) \tag{10.12}$$

Thus taking only into account these two variables A and B, for 2 binding sites, there are 6 possible mechanisms and as many as 27 for a 3 sites interaction. ITC can be used to distinguish between these possibilities only for a small subset of mechanisms, since different mechanisms can lead to non distinguishable binding curves. As a result fitting of a site-model to the data does not guaranty the appropriateness of the model and model comparison strategies are required.

10.2.5.3 Strategy for ITC data analysis

Criteria for model selection Because of the multiple models fittable to ITC data, a reliable way to compare models is required. The basics and the choice of this comparison method are explained in this paragraph. Let's consider the comparison of successive macroscopic models of stoichiometry 1 and 2. The two candidate models form what is called nested models. The 2:1 model has four thermodynamic parameters: two dissociation constants and two enthalpies for the two successive binding reactions. The 1:1 model is a particular case of the 2:1 model where the second binding constant and enthalpy are equal to zero. This is why it is said to be nested into the more complex model. In the case of nested models, the model with the largest number of parameter (*i.e.* 2:1 here, 4 thermodynamic parameters) will always fit at least as well - and usually much better - than a model with less parameter (1:1 model, 2 thermodynamic parameters). As a consequence the quality of the fit, estimated for instance by the RMSD value, will be always better for the model with the highest number of parameters than for the smaller nested models. However this does not mean the more complex model is the more appropriate. The challenge in model selection is to take this phenomenon into account to be able to determine for a dataset and various candidate models, which one is best supported by the data. To do this, it is necessary to take into account the fit quality and the number of parameter of each model. Various indicators are commonly used for model selection. They include reduced Chi-square statistic, Fisher test and various information criteria, in particular the Akaike Information Criterion (AIC). The reduced Chi-square requires the knowledge of the number of degree of freedom that can be determined only for linear models, not used in ITC. Akaike Information Criterion has theoretical advantages over the Fisher test strategy. It is has been shown in various works (101; 145; 177) in particular using Monte Carlo simulations that the Akaike criterion is more reliable than the Fisher test method for different type of models and datasets. Moreover the AIC can be applied to compare both nested and non nested models as the ones of step 1 and 3 the proposed fitting strategy. Finally, the Akaike criterion enables to compare simultaneously multiple models, it is not restrained to binary comparisons. For these reasons AIC has been used in this work. A comprehensive review of the theoretical and practical aspect of Akaike Information Criterion can be found elsewhere (24)

The nature of Akaike Information Criterion (AIC), the practical calculation of Akaike weights and its interpretation are developed in this paragraph. The Akaike Information Criterion has been proposed by Hirotugu Akaike in 1973 (4). The AIC value is an estimation of the relative distance between the fitted model and the unknown mechanism that generated the data.

Let's consider n data points and a set of candidate models i with k_i parameters. We consider here the particular case, used in this work, of model parameters estimated using maximum likelihood strategy with normally distributed errors and a constant variance. In this situation, the residual sum of square (RSS) can be directly used to calculate the Akaike Information Criterion (AIC) value for each model (24):

$$AIC_i = n\, ln\left(\frac{RSS}{n}\right) + 2k_i \tag{10.13}$$

The deviance of the model $n * log(\frac{RSS}{n})$ is penalized by two times the number of parameters (k_i). The demonstration of this result is out the scope of the present explanation and can be found elsewhere (24). Because the number of points in a traditional ITC experiment is between about 15 to 40 points (31 and 39 here but actually more due to global fitting of multiple dataset) and that the number of parameters to fit is a significant proportion of the number of data points, the AIC value has to be corrected for this "small sample size" effect. The necessity of this correction has led to the introduction of the reduced AIC value by Hurvich and Tsai (129).

$$AICc_i = AIC_i + \frac{2k_i(k_i+1)}{n-k_i-1} \tag{10.14}$$

This formula is widely considered to be the standard way to calculate Akaike value since the limit of the fraction when n tend to $+\infty$ is 0. Because of AIC definition, the model with the lowest $AICc$ value (if any) is the preferred one. It is the differences between $AICc$ values that are of interest to compare the other models to the preferred one.

$$\Delta i = AICc_i - AICc_{min} \tag{10.15}$$

It is often difficult to judge from the AICc differences if one model is slightly or really worse than the preferred one. Although some ranges of AIC differences have been proposed to qualify the model considered, a unified and clear approach consists in the calculation of the Akaike weights that correspond to easily interpretable probabilities.

For each model i among a set of M candidate models, the Akaike weight is the ratio of the relative likelihood of the model i over the sum of the relative likelihoods over the different models :

$$w_i = \frac{exp(-\frac{1}{2}\Delta i)}{\sum_{j=1}^{M} exp(-\frac{1}{2}\Delta j)} \tag{10.16}$$

The relative likelihood of the model i given the data, w_i, can be interpreted as the probability that the model i is the preferred one for the data and the considered set of candidate models. It is important to underline that these weights enable to compare models but do not indicate that any of them adjust well the data. These probabilities are called weights because when various models are well supported by the data, the fitting can be performed using these multiples models weighted by these probabilities.

Proposed general strategy to analyze ITC datasets I propose the following general strategy to fit the ITC datasets. The importance of fitting first macroscopic models has already been underlined in (86). A general strategy using global fitting of the ITC datasets followed by model comparison with Akaike weights is presented here. This procedure is different from a common one consisting in starting from the simplest models to the more complicated one and to stop when a good fit is obtained which can lead to an over/mis interpretation of the ITC data.

(1) Global fitting of the data on different macroscopic models of different stoichiometries. For a model of molar ratio n, the number of parameters in the macroscopic model is $2n$. Akaike weights are used to select one (or more) models (i.e. the molar ratio) among the ones that can be fitted. This first step is valid whatever the microscopic mechanisms for ML_n models providing the data enable reliable fitting of the $2n$ variables.

(2) Depending on the datasets and potential additional information, one can try to fit the datasets with site-specific models. If this step is to be used to get insight of the microscopic mechanisms, all possible models should be checked with and without the selected macroscopic model of the previous step. The comparison of microscopic models might reduce the number of possible mechanisms. Some of these microscopic models have less variables than the macroscopic one (e.g. independent identical sites) and could be the only model fittable to the dataset but in such case the hypotheses associated to the model has to be explicited and discussed.

10.3 Development of an open source software for the automated analysis of ITC data

10.4 R-ITC experimental design and data analysis

Despite some ITC program are free of charge, no open source program able to fully process the raw ITC data have been released by the time of writing. I consider this to be necessary so that the users can actually understand in detail what they are doing in each step of the ITC processing having access to a fully commented code. Also many programs do not facilitate multiple model comparison and global fitting. Moreover there is a clear tendency to speed up the ITC measurement, by reducing the sample size and increasing the sensitivity and to provide high throughput calorimeter. As a consequence, more dataset are generated and the usual programs have not been conceived to do batch analysis of multiple datasets. Many of the available programs require user input or adjustments that are not only time consuming but are also a source of variability and bias in the data analysis. A comparison of the main available programs is given in the following table.

Table 10.2: Comparison of the main softwares for Isothermal Titration Calorimetry data analysis. [1] Sedphat (126) ; [2]: suITCcase (web form (http://alanine.usc.es/suITCase/suITCase.php)

name	raw data	source	extensible	automated	simulations
Origin + ITC module	yes	no	yes	no	1:1
Sedphat[1]	no	no	no	no	no
suITCcase[2]	no	no	no	no	1:n

Also some in house programs have been written, in Pascal, C, Matlab®, Mathematica®, Scientist®. They have not been made available nor described.

I've written a program for ITC data fitting called RITC to be released latter as an R package (free, open source) . GNU R (228) is a widely used language that has become the gold standard for many biological data analysis. The latter program implements all the previously defined objectives and enable to do multiple processing of ITC data.

10.5 Non trivial indiscernible processes in ITC for the study of equilibrium

- {Reaction signal masked by other co-reactions or by the noise}, {reaction conditions (especially temperature) where the binding enthalpy is almost zero} or {no reaction happening} can all lead to no significant signal detected during the mixing. No signal detected during the mixing does not imply that no reaction is happening. Experiment at different temperatures with at least 10°C and / or experiments with higher concentrations can help to find the existence of a reaction.

- For any non null integer n, {if the M and L species exit only in solution and in the M/L complexes as M_m and L_m, leading to complexes of type $M_mL_{m.j_{j\in 1:n}}$ } or { if the M and L species exit only in solution as M and L and the complexes between M and L are only of type $M_mL_{m.j_{j\in 1:n}}$ } then the molar ratio of these situation is the same than the reaction leading to a ML_n complex by progressive binding of L on M. This ratio is n and is not the stoichiometries of the two complex situation mentioned. This is why the notion of "molar ratio" is systematically used in this work rather than the stoichiometry, often used in the literature. Other methods or information are required to resolve the actual stoichiometry.
- In the working conditions, a system with a non empty set of reactions with negligible enthalpy change, associated to a non empty set of reactions with significant enthalpy change can lead to a similar signal model than a system possessing only the set of reactions with significant enthalpy change. This is why the molar ratio is qualified in this work to be "apparent". The term apparent stoichiometry is sometimes used in the literature and in UIPAC recommendations for ITC (244). Care is required in the interpretation of these key parameters that appear as minimums.
- The potential reactions including the one to be studied, co-reaction buffer protonation, reaction related to buffer mismatched and oligomers dilution reaction can all lead to a titration signal. Nothing indicates the contribution of the reaction under study in the ITC signal. In order to get relevant thermodynamic parameters, the experiment has to be design so that all the other processes are negligible or can be estimated with multiple control experiments.
- Let's consider an ITC experiment and the repetition of this experiment performed by varying one parameter, for instance the addition of one ligand. If {the parameter has an effect on an undetected reaction} or if {it has no effect at all}, no difference will be seen on the ITC signal. So it is not always possible to directly interpret one negative ITC experiment performed in one condition.

Chapter 11

svAUC: protocol for the study of reversible interactions and implementation

Definition and main biophysical applications of svAUC In Sedimentation Velocity Analytical Ultra Centrifugation (svAUC) method, some time dependent optic signals profiles in sector shaped cells are monitored during a centrifugation performed at a high constant acceleration (usually 100000-300000 $\times g$) in an analytical ultracentrifuge. Common applications of svAUC include the purity analysis of a macromolecular sample with almost no restriction of the size limit, the stoichiometries of complexes, the overall shape of complex and species in solution, the dissociation constants and the thermodynamic non ideality of the solution.

11.1 Principle and implementation of svAUC

1. Centrifuge, rotor and cells - implementation Sedimentation velocity experiments were performed with a Beckmann Coulter XLI analytical ultracentrifuge with an An-50Ti rotor. The rotor is equipped with 8 holes. One of them is filled with a counterbalance. The 7 remaining holes can contain double cells that can be filled with samples and buffers. The individual cell are sector shaped (**fig. 11.1**) to avoid convection. Contrary to a preparative one, an analytical centrifuge can measure the distribution of some signal in the sample during the spinning. The cells were sapphire-windowed cells with 12 mm optical paths. Sapphire window enable the UV light to pass through the cells and are highly resistant to a wide range of physico-chemical conditions (pH, temperature).

2. Preparation and filling of the centrifuge cells The elements constituting the cells require an extensive wash before their filling. Then the reference cell is filled with the buffer alone and the sample cell with the sample solubilized in the same buffer. Some buffer component might distribute slightly differently during the ultracentrifugation and the reference cell enable to take into account a resulting potential change in the absorbance profile of the solvent.

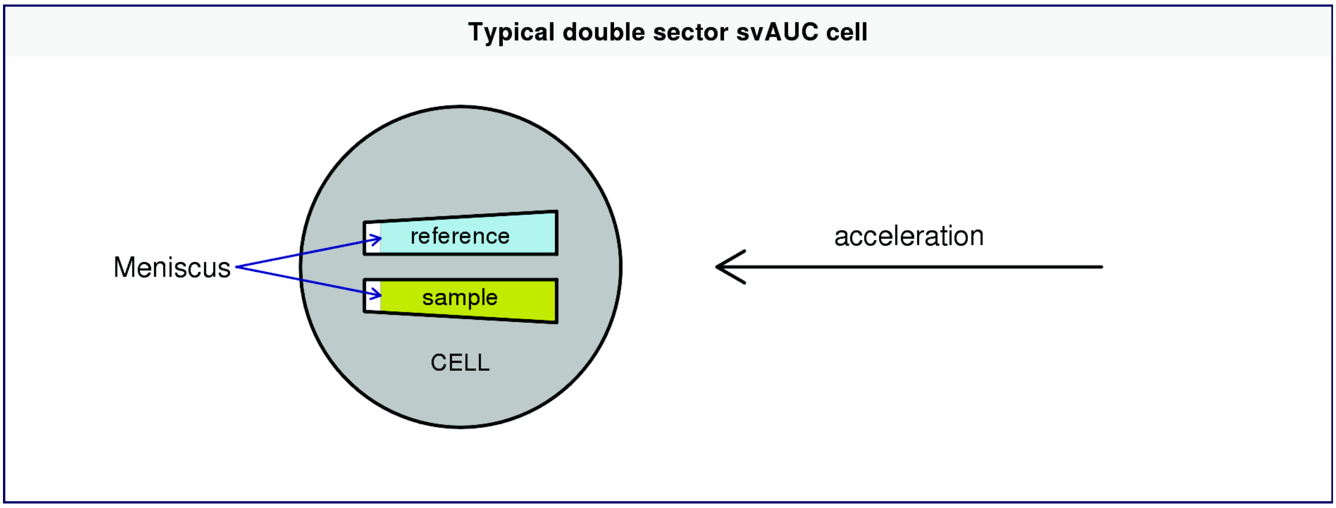


Figure 11.1: Typical configuration of a svAUC cell with two sectors shaped cells, one for the reference buffer and the other one for the sample to be studied. The cells are not fully filled and the cell with the reference buffer is slightly more filled than sample cell to detect a well defined sample meniscus in the scans. The windows, not visible in this top view, are discs placed on each planar side of the cell.

Implementation. The cells elements were washed with detergent then extensively washed with distilled water and treated 5 min with RNaseZAP (Sigma-Aldrich). An additional long wash with RNase free water was performed to remove potential traces of RNaseZAP (Sigma-Aldrich), and the cell were allowed to dry. Assembly of the cells under RNase free conditions enabled to get RNase free cells. The cells were filled using a Hamilton syringe that received the same treatment that the cells. The buffer used, as for the other biophysical techniques is: 23.5 mM potassium phosphate, 100 mM KCl, 10 mM $MgCl_2$ and 5 mM β-mercaptoethanol

3. Equilibration of the centrifuge - implementation Before the experiment, the centrifuge was equilibrated at 25°C during 40 minutes. During this step a slow speed was applied to check the waterproofing of the cells before starting the experiment at full speed.

4. Centrifugation and cell scanning During the centrifugation in an analytical ultracentrifuge, the cells undergo a constant high centrifugation speed usually between 100000- and 300000 $\times g$. This centrifuge works under approximate vacuum to avoid frictional heating and turbulence. The concentration distribution of macromolecules across the cell at various times during the experiment is measured while the sample is spinning, using either absorbance, refractive index or fluorescence detection. During the centrifugation, the typical Beckmann XL ultracentrifuge detection systems monitor the redistribution of molecules in the cells with two detection systems: UV absorption optics and Rayleigh interference optics. The former system performs a scan along the cell axis and is useful for macromolecules absorbing significantly in UV like protein and nucleic acid. The concentration dependence of the UV signal is the well known Beer-Lambert law. The Rayleigh interference is based on the split of a monochromatic coherent light beam that passes through two slits placed below the two pairs of cells. The beams are eventually merged causing interference fringes pattern. If the refractive indexes in the sample cell and in the reference cell are different, the optical pathes have not the same length and it creates an interference figure. If the refractive index of one the two pairs cells change in one region on the centrifugal strength axis, because of the presence of a macromolecule for instance, there will be an orthogonal shift of the interference fringes directly proportional to the change of concentration so the knowledge of the concentration at one point can enable to determine the concentration at any other point.

Implementation. The centrifugation was performed at 50,000 rpm (185,000 g) at 25°C during 15 hours. Acquisition was performed for UV absorbance at 260 nm, 280 nm and using interference optics with a laser at 655 nm. For the UV signal, the buffer subtraction is performed by the recording device, the scans of the cells were recorded at $\approx$ 10 minute intervals with alternating scans at 280 nm and 260 nm from 5.85 cm to 7.25 cm to the rotor center, each scan being the average of two replicates. Interference scans were recorded every 20 minutes.

11.2 svAUC data analysis, approximations and related biases

11.2.1 Transport processes in AUC

The svAUC cells are sector shaped with a carefully chosen geometry in order to avoid convection of the particles during the centrifugation. The design of cell to avoid this phenomenon is one the most important reason of the applicability of svAUC. From here, it is considered that no convection occur in the svAUC cells ($\mathcal{A}_{AUC}1$). Despite biological macromolecules are usually charged, the presence of sufficient salts in the environment compensate these charges so it is reasonable to approximate that the particles transport behaves like uncharged ones. The following development is valid for uncharged particles or particle with a similar behavior $\mathcal{A}_{AUC}2$. Under this hypothesis, two transport processes take place during the ultracentrifugation: sedimentation and diffusion

Sedimentation corresponds to the "separation of a dispersed system under the action of a gravitational or centrifugal field according to the different densities of the components" (134). Sedimentation depends on the relative density of the particle compared the environment one and on the friction of the particle in its environment. If the density of the macromolecules is higher than the one of the solvent, the high centrifugation depletes progressively the macromolecules from the meniscus, forming a boundary that moves to the bottom of the cell until the centrifugation is stopped or until all the macromolecules form a pellet. For entities with density lower than the solvent, the sedimentation transport happens in the opposite direction. The sedimentation of a particle under the acceleration a causing the sedimentation at a speed v is characterized by its sedimentation coefficient :

$$s = \frac{v(t)}{a} \tag{11.1}$$

In order to get simple figures for biological macromolecules, the sedimentation coefficient is usually expressed in Svedbergs, with $1S = 10^{-13}s$ but values in seconds have to be used in svAUC equations. The sedimentation value is intensive and depends mainly on the properties of the particle and on its environment (buffer, temperature, etc). In non ideal mixtures, the concentration of the particle itself can also increase the friction of the particles compared to an infinite dilution situation. This causes a decrease of the sedimentation coefficient. Let's consider a particle at infinite dilution and let s and f be the sedimentation and frictional coefficient respectively, s_c and f_c the sedimentation and frictional coefficient respectively corrected for non ideality and k_s an experimental correction constant in ml.g^{-1}. A first order correction is usually sufficient to model the effect of the concentration on the sedimentation coefficient (118) :

$$f_c = f(1 - k_s c) \tag{11.2}$$

leading to :

$$s_c = s(1 - k_s c) \tag{11.3}$$

For solution behaving very closely to ideal ones, we have $(1 - k_s c) \approx 1$ ($\mathcal{A}_{AUC}3$). This hypothesis becomes wrong in the bottom of the cell where the concentration of macromolecule are increasingly high during the centrifugation. As a result this region is also characterized by a much larger contribution of diffusion than is the rest of the cell because of the large concentration gradient and almost no sedimentation (bottom of the cell). Thus, the data acquired at the bottom of the cell region are not included in the type of analysis described here. The use of $\mathcal{A}_{AUC}3$ is common in AUC and appear very useful when considering fitting of svAUC equations.

Diffusion characterizes the process by which macromolecules are transported as a result of Brownian motion in the absence of turbulent mixing. The driving process of diffusion is the gradient of chemical potential (118). During the ultracentrifugation, diffusion in the cells tends to equilibrate the increasingly different chemical potentials (including differences of concentrations) of macromolecules. The diffusion process in ultracentrifugation is monodimensionnal and characterized by the translational diffusion coefficient D (in $m^2.s^{-1}$). Like the sedimentation coefficient, the diffusion coefficient depends on the concentration of the particle but in a slightly more complicated manner than the sedimentation coefficient. A derivation of the relation between D_c and c from first principles has been reported (118) and it can be well approximated by a linear formula, with D the translational diffusion coefficient at infinite dilution :

$$D_c = D(1 + k_D(s)c) \tag{11.4}$$

Very often in AUC, the $k_D(s)c$ term is considered negligible compared to 1 here ($\mathcal{A}_{AUC}4$), which require working at very low concentrations.

11.2.2 Relation between the transport processes and the concentration gradient: the Lamm equation.

For uncharged particles of any shape in solution that undergo only monodimensionnal diffusion and sedimentation transport processes in a sector shaped cell, the concentration law for each species can be obtained from the Fick law and the continuity equation, leading to the Lamm equation, with r the distance from the axis of rotation, t the time from the start of the centrifugation experiment, ω the rotor angular velocity, i one of the n species participating in the reaction and q_i the chemical reaction flux of i (241) :

$$\forall\, i \in [\![1, n]\!], \quad \left(\frac{\partial c}{\partial t}\right) + \frac{1}{r}\frac{\partial}{\partial r}\left[r\left(s_i\omega^2 r c_i - D_i\frac{\partial c_i}{\partial r}\right)\right] = q_i \tag{11.5}$$

q_i equal zero for non interacting species. Once the number of species in solution is known, this system of equation can be solved numerically, potentially using the different kind of signals sensitive to the concentration. This is usually performed after a $c(s)$ analysis described below than can be use to determine whether the system behaves like a non interacting system; if not knowledge of the reaction parameters (dissociation constants and kinetics constants) is required to model the q_i terms. When the sedimentation and diffusion coefficients depend on the unknown concentration, the Lamm equation becomes nonlinear and hence more difficult to solve. This is not the case here because of the previous approximations ($\mathcal{A}_{AUC}2, 3$)

Implementation Lamm equation modeling with multi-wavelength analysis was performed with the program Sedphat (10).

11.2.3 Svedberg equation

The Svedberg equations are independent to the Lamm equation and give a relation between the diffusion and sedimentation coefficients. It enables extraction of mass and shape information from the knowledge of the ratio of the sedimentation coefficient and the diffusion coefficient. The historical formula which is still the main one used in svAUC can be derived with the second law of motion for a 2 component uncharged system, a molecule i in an organic solvent (171) and assuming $D = \frac{RT}{N_A \cdot f}$ with R the gas constant, T the temperature, N_A the Avogadro constant, f_i the frictional coefficient of i, $\overline{v}$ the partial specific volume of i (in ml.g^{-1}) and ω the rotor angular velocity (in rad.s^{-1}) :

$$\frac{s_i}{D_i} = \frac{M_i(1-\rho\overline{v}_i - s_i^2\omega^2)}{RT} \tag{11.6}$$

with the term $s_i^2\omega^2$ being negligible and usually not reported in the equation. $M_b = M(1-\rho\overline{v})$ is called the buoyant mass. However in svAUC of biological polymers, the system is usually not just a two component system. This is usually not explained in AUC textbook or articles, that usually start with the historical Svelberg equation without reminding its conditions of validity. This has been the subject an interesting vigorous article-interposed debate in *Protein Science*: "Modern analytical ultracentrifugation in protein science: a tutorial review" (160) criticized in "Modern analytical ultracentrifugation in protein science: look forward not back" (67) and answered in "Back to the future: A rebuttal to Henryk Eisenberg" (161). The key question that underpins these articles is whether it is more useful/rigorous to start from the approximate historical equation or the more rigorous equation derived from a multicomponent theory, both requiring very low concentration :

$$\frac{s_i}{D_i} = \left(\frac{\partial\rho}{\partial c_i}\right)\frac{M_i}{RT} \tag{11.7}$$

A direct use of last formula would require an estimation of $\frac{dp}{dc}$ using density estimations. The main limitation is that the measurement of $\frac{dp}{dc}$ with standard densitometer currently available (e.g. Paar DAM 5000) is the large amount of material required, 1 ml at 3 to 20 mg.ml^{-1} respectively. Differences between the resulting estimated mass terms have been calculated in these articles from system that have been studied in depth. For diluted protein solution the estimation of these parameter on a 1 mg.ml^{-1} solution without concentrated cosolvents is susceptible to add an error larger on the molecular mass ($\approx$1.3%) than the effect on the correction on this parameters ($\ll$ 1%). Anyhow, the choice of the approach might just be constrained by the available material. Because of the limited material available in this study and of the handling risk to manipulate RNA in potentially RNA contaminated devices, the common approach using the historical Svelberg equations has been used. The main point of this paragraph is to underline that because of the resulting approximation and according to the experimental data provided in the series of article discussing this issue (160; 67; 161), a bias up to 9 % might affect the estimation of the molecular weight of RNA and protein/RNA complexes.

Practical use of the classical Svedberg equation The classical Svedberg equation is usually applied by introducing other parameters closer to experimental information. With $D = \frac{RT}{\mathrm{N_A} \cdot f}$, with the Stockes equation for a sphere in a solvent of viscosity η and radius R_0 ($f_0 = 6\pi\eta R_0$), with R_0 expressed as (160) $\left(\frac{3M\overline{v}}{4\pi \mathrm{N_A}}\right)^{\frac{1}{3}}$ and by replacing R_0 by the hydrodynamic radius $\frac{f}{f_0}R_0$ to account for the lack of sphericity, the sedimentation coefficient of a species i can be expressed as :

$$s_i = \frac{M_i(1-\rho\overline{v}_i)}{\mathrm{N_A} 6\pi\eta \frac{f_i}{f_0}\left(\frac{3M_i\overline{v}}{4\pi \mathrm{N_A}}\right)^{\frac{1}{3}}} \tag{11.8}$$

This equation enables for instance to predict the sedimentation coefficient of a macromolecule (often used in this study). The ratio of the frictional coefficient of the macromolecule and the one of an equivalent sphere in the same solvent, $\frac{f}{f_0}$, is called the frictional ratio. The larger the frictional ratio, the higher the difference between the sedimentation of the particle under study and the one of a sphere, which sediments faster. Experiments have shown that globular proteins have a frictional ratio of about 1.2 whereas elongated particles has higher frictional ratio, which can be over 2.

11.2.4 $c(s)$ analysis of the svAUC signal

Principles and implementation One of the first common methods to analyze the svAUC signal is to determine the diffusion deconvoluted sedimentation coefficient distribution $c(s)$, which is defined by the following integral equation, with $d(r,t)$ the experimental svAUC data, $c(s)$ the concentration of species with sedimentation coefficient between s and $s+ds$, $\chi(s, D(s), r, t)$ a (discrete numeric) solution of the Lamm equation and ϵ the noise :

$$d(r,t) = \int_{s_{\min}}^{s_{\max}} c(s)\chi(s, D(s), r, t)ds + \epsilon \tag{11.9}$$

The numerical Lamm solution used in this equation correspond to a system of a large number of discrete non interacting species (46). This approach requires an estimation of the diffusion coefficient of the species involved from their estimated (inputs) partial specific volume and frictional ratios using the previously described Svedberg equation. $c(s)$ profiles are thus appropriate for non reacting mixture but can be also applied with a good approximation to slowly or rapidly exchanging systems (next paragraph). The result resembles a SEC profile.

Implementation The $c(s)$ analysis was performed with SEDFIT Version 12.52 (241).

Application of the $c(s)$ analysis to protein/RNA complexes Protocols for svAUC analysis of protein-RNA complexes with the $c(s)$ analysis have been described (282) and svAUC analysis of associating systems was recently reviewed (39). The application of $c(s)$ analysis to interacting systems is an efficient way to describe the main sedimentation properties of these systems (46; 243; 242; 303). Three kinds of behavior can be detected in the $c(s)$ analysis of a complex depending on the kinetics of the experiment. In an experiment with a concentration variation of one of the two components, the other being in excess, an undisturbed peak is expected for the compound in excess. Additionally, if there is n possible complexes, up to n peaks can be detected. The sedimentation characteristics of these peaks called reaction boundaries depends on the sedimentation properties and kinetic of the complex *and* on the free species (303). All these time scale are very different from the NMR ones and these regimes are all slow for NMR.

- *Slow exchange* For complexes with half-life over about 10^4 seconds the interaction is slow on the time scale of svAUC and the properties of these boundaries corresponds to the one of the individual complexes. The contribution of the free species is negligible. In this situation, the integration of the $c(s)$ profile directly gives an estimate of the concentration of the corresponding species. Binding constant can be determined from a series of such distributions.
- *Fast exchange* For complexes with half-life less than 10^3 seconds, the s-values and buoyant masses of the concentration boundaries are concentration dependent and thus do not give direct information of the complexes formed. For a system with n components, there will be n boundaries. The integral of this reaction boundary/ies is obviously not related to the amount of a given complex. The s-value of the reaction boundary will depends on the s-value of all the other species, their concentration and on the equilibrium existing between them. The interpretation is thus much more complex than the slow exchange case. However this is a very common case.
- *Intermediate exchange* Between the fast and intermediate regime, a mixture of the two extreme cases happen.

11.2.5 Determination of the parameters used in the svAUC analysis

The use of the previous equations and the fit of related models require some knowledge :

- The mass density (1.005 g.ml^{-1}) and viscosity (0.898 cP) of the phosphate buffer used were calculated using Sednterp (v20120328 beta, http://bitcwiki.sr.unh.edu/index.php, (157)).
- The molar masses for the polymers were calculated from their composition.
- The partial specific volume ($\overline{v}$) can be determined from the known composition of the macromolecule. It is however dependent on the solvation, which has led to the introduction of the "partial specific volume of the macromolecule v", a property dependent only on its composition and the "effective partial specific volume" v_e, relevant in solution. The ratio v_e/v has been determined in some studies and it appears to be significant (118). It reaches 5 for human macroglobulin. A possibility, which has been used in this study, is to use experimental values determined for a similar system in similar conditions. This was particularly desirable for the RNA for which few data is available compared to proteins. The partial specific volume used for pre-miR-155 (0.508 ml.g^{-1}) was taken from a study of a viral RNA in dilute aqueous KCl solutions (20). A partial specific volume of 0.74 ml.g^{-1} was estimated for TRBP-D1D2 using the software Sednterp version 2 (http://bitcwiki.sr.unh.edu/index.php,(157)). The partial specific volume for possible protein/RNA complexes was estimated by summing the individual values of each component weighted by their fractional mass in the complex.

Chapter 12

Introduction to NMR spectroscopy to study secondary structures and some exchange processes

12.1 Principles of liquid state NMR spectroscopy

Definition and main biophysical applications of NMR Nuclear Magnetic Resonance spectroscopy (NMR) is the spectroscopy of the atomic nuclear Zeeman sublevels i.e. the ones within the nuclear ground state. Compared to other spectroscopic techniques the quantum states maintain very long coherence and these states are very diversely controllable. The liquid state version of NMR, referring to the sample state, enable to get structural and dynamic information on molecules in liquids in their apo form or in complexes with a site specific information at the atomic nucleus level. It is to date the only technique enabling to study at atomic resolution molecules in liquid state. NMR data can also be used to study kinetics and thermodynamic parameters. NMR is the main technique used in this work to get site specific and dynamical information on the TRBP/miR-155 system that the other methods couldn't solve. The method was mainly applied to get information on the secondary structure and flexibility of macromolecules and for the study of chemical exchange process (interactions and conformational exchange). This introduction focuses on these topics.

12.1.1 Liquid state NMR samples

Common easily NMR detectable nuclear isotopes in biological macromolecules Only the isotopes with non zero nuclear spin can produce NMR signals. For biological molecules, the main interesting nuclei for NMR and properties of the defavorable abundant ^{12}C and ^{14}N are given for comparison in **tab. 12.1**(194)

NMR sample and isotopic enrichment For proteins and nucleic acids, the nucleus H, N and C are very abundant and widespread within the polymers chain. This makes them attractive

Table 12.1: Main isotope of interest for NMR spectroscopy of biological macromolecules as well as abundant ^{12}C and ^{14}N given for comparison (*: unstable isotope)

Isotope	**Ground state spin**	**Natural abundance (%)**	**Gyromagnetic ratio (10^6 rad.s^{-1}T^{-1})**
^{1}H	1/2	$\approx$ 100	267.522
^{2}H	1	0.015	41.066
^{3}H*	1/2	0	285.349
^{12}C	0	98.9	NA
^{13}C	1/2	1.1	67.283
^{14}N	1	99.6	19.338
^{15}N	1/2	0.37	-27.126
^{17}O	5/2	0.04	-36.281
^{19}F	1/2	$\approx$ 100	251.815
^{31}P	1/2	$\approx$ 100	108.394

for NMR especially because of the advantageous spin $\frac{1}{2}$ of the proton and of two rare isotopes of carbon (^{13}C) and nitrogen (^{15}N). The samples for NMR are often prepared using isotopically enriched precursors in order to get macromolecules that are fully or site-specific isotopically enriched with NMR favorable isotopes.

Implementation The production and enrichment protocols used for the preparation of NMR samples for the work reported here are given in part III. The large majority of the NMR experiments performed in this thesis were done using concentric tubes as represented and commented in **fig. 12.1**.

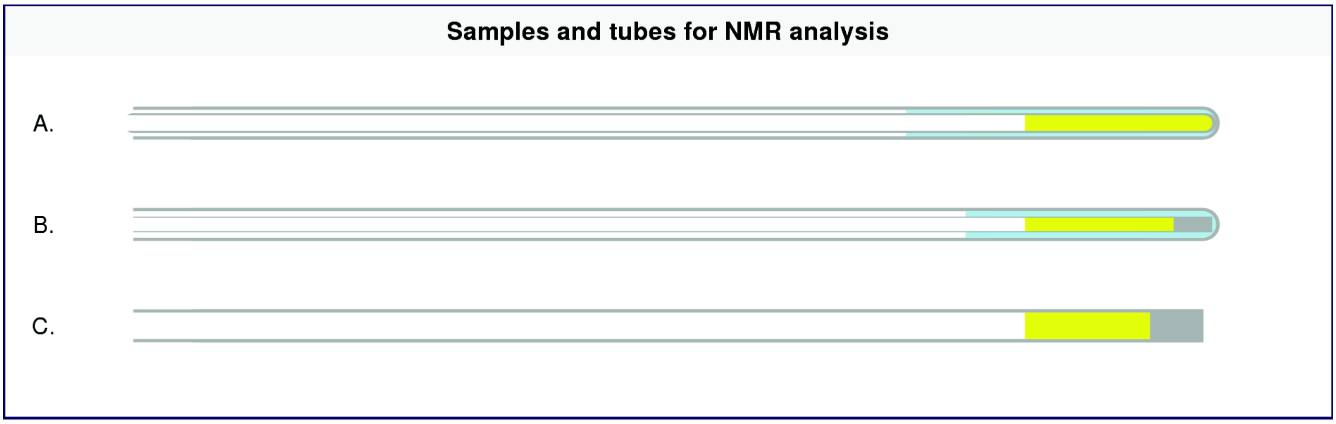


Figure 12.1: Representation of the three different configurations of NMR tubes used in the work. Samples are represented in yellow and heavy water in light blue. Tubes of 4.2 mm inner diameter (Norell®), 5 mm outer diameter) are filled with 300 μl $^{2}H_2O$ and either with a thinner tube of 2.4 mm inner diameter tube (Norell®), 3 mm outer diameter) containing about 120 μl sample without added $^{2}H_2O$ (**A.**) either with a 1.8 mm inner diameter Shigemi tube (Shigemi Inc.) with about 55 μl sample without added $^{2}H_2O$ (**B.**). When more material and/or sensitivity were required a larger Shigemi tube of 5 mm outer diameter (Shigemi Inc.) with 250 μl to 300 μl of sample with 5% $^{2}H_2O$ was used (**C.**). Shegemi tubes were used without plungers in the titrations to avoid material loss and biased protein:RNA ratios.

12.1.2 NMR spectrometer

Three different NMR spectrometers were used in this study, with DirectDrive consoles :
- 800 MHz (18.8 T) with a cryogenic (^{1}H,^{15}N,^{13}C,^{2}H).
- 600 MHz (14.1 T) with a cryogenic (^{1}H,^{15}N,^{13}C,^{2}H).
- 600 MHz (14.1 T) with a room temperature probe (^{1}H,^{15}N,^{13}C,^{2}H,^{31}P).

As all spectrometers based on Fourier transform spectroscopy, they are composed by :
- a temperature controlled zone where the sample is placed, enabling to set the desired temperature for the experiment.
- a superconducting coil generating an intense and highly homogeneous B_0 field (T). Because of the extremely homogeneous field required for NMR, once the sample is in the magnet, the electric current passing through a set of additional coils (about 40) called shims and positioned at different places around the sample, is optimized to get a field as homogeneous as possible (shimming procedure). For NMR performed on earth, the actual magnetic field felt by the sample is the sum of B_0 more the radiofrequency field from the transmitter more the potentially fluctuating environmental magnetic field, essentially the earth field of about 0.05 mT, which is completely negligible compared to B_0.
- a probe containing a coil called transmitter that both generate the radiofrequency pulses and record in first the radiofrequency signal. Since most of the NMR noise comes from the thermal noise in the coil, the sensitivity can be enhanced (up to 4 times) by using probes with coil temperature of a few Kelvin (cryoprobes). Two capacitor capacitance are adjusted before each experiment so that the resonance frequency of the electrical circuit including the coil matches to the one of the frequency of the spectrometer (tuning procedure).
- a pulse generator, composed of a series of a computer controllable frequency synthesizer, a controlled gate that will defined the incoming radio-frequency pulse, a tunable attenuator, and a constant amplifier connected to the transmitter.
- two low noise NMR signal amplifier in series.
- an analog to digital converter that digitize the NMR signal for computer storage.

12.1.3 Principles of NMR experiments

One of the goals of NMR experiments is to get spectra with peaks corresponding to the frequency of allowed Zeeman sublevels transitions, of shape as close as possible to absorption peaks and from which can be extracted structural and dynamic information . This is however not straightforward. Introductory quantitative quantum mechanical description of the origin of the NMR signal can be found in many NMR textbook including the ones of M. Levitt (194) and J. Keeler (130) for the foundations and J. Cavanagh (29) especially for the description of many biological NMR experiments.

Principles of NMR experiments NMR experiments manipulate the spin states using the properties of different Hamiltonians in order to get information on the related interactions. During Fourier Transformed NMR experiments, the initial magnetization created by the permanent field B_0 is altered by applying temporarily short radiofrequency pulses creating a weaker magnetic field in a plane orthogonal to the B_0 direction. This additional magnetization will determine the nature of the experiment performed and the goals are not detailed here. Example of the EXSY experiment is given below. These fields can excite at the same time a wide range of frequencies. The first pulses rotate the bulk magnetic moment towards the transverse plane, leading to a non null transverse magnetization and a reduced component along the B_0 direction. After the application of these magnetic fields, the exited spins angular momentum will undergo precession. Application of multiple radiofrequency pulses in series with delays determines the experiment performed. For a given nuclei type, the frequency of the precession of each individual atoms will be influenced by the magnetic perturbations caused by the chemical environment. This environment corresponds to specific electron density and fields that depends on the nature of the surrounding atoms, their spatial structure, bounding network and dynamics. As a result each spin is characterized by a frequency (Larmor frequency) that behaves like a probe of the chemical environment. Each spin contribute to build a signal that is recorded over time by monitoring the resulting electric current that pass through the coil that was used for the excitation. This time dependent signal is the fundamental observable in NMR and called the free induction decay (FID). The FID loses exponentially coherence over time because of the relaxation of the system. It is then converted to a frequency spectrum using the Fourier transformation giving Lorentzian peaks characterized by their amplitudes, frequencies (given in chemical shift to account for the field dependency) and linewidth. These are the secondary NMR observables.

Implementation All experiments were run with the VNMRJ (v3) program (Varian).

NMR Signal processing - implementation All the NMR spectra were processed by NMRPipe (51) with the classical following steps :
- Zero filling of the free induction decay: As the fast Fourier transform used in NMR processing (usually the common Cooley–Tukey algorithm) is faster for power of two number of points, the signal matrix can be completed by zeros in order to reach a more appropriate size. Increasing more the matrix with zeros (usually 2 times in this work) enables to have better defined NMR peaks (more points) but does not change the resolution.
- Apodization: During the acquisition time, the maximum of the signal decays exponentially, so the maximum occur at the beginning and then the signal to noise ratio tend to zero. To enhance

the sensitivity by weighting more the first portion of the signal and to enhance the resolution by weighting less the end of the signal, a so called apodization function is multiplied to the signal. In this work, a square sinebell function was used. The resulting NMR peaks are thus no longer Lorentzian but closer to a more compact Gaussian distribution with a much narrower base than a Lorentzian distribution.
- Fast Fourier transform: when the sampling of a multidimensional (nD) spectrum is uniform, the exact transformation to go from the time dimension to a frequency dimension is given by an nD Fourier transform.
- Phase correction approximation: The phase of the amplitude of the signal components depends of the experiment. For the experiments used in this work, the phase of the peaks depends approximately linearly on the center of the spectral region. Thus the phase can be approximately corrected for all the peaks by finding a linear correction of the phase on each dimension in order to reduce dispersion modes in the spectrum.

12.2 Analysis of the secondary structure by chemical shifts and NOE

TRBP constructs secondary structure analysis Protein backbone chemical shifts and especially ^{1}Hα, ^{13}Cα and ^{13}Cβ depend on the local backbone conformation (281). Backbone resonances of proteins are commonly assigned to their corresponding atoms in the sequence with 3D HNC correlation experiments. This information enables to identify regions of stable secondary structures or secondary structures propensities using empirical rules obtained from proteins of known structures. In this work, secondary structures were predicted using TALOS+ (247), an algorithm based on a database of structure of known high resolution x-ray structure and NMR chemical shift data. These predictions were complemented with heteronuclear ^{1}H-^{15}N Nuclear Overhauser Enhancement (NOE) experiments which provide information about the motions of ^{1}H-^{15}N bond vectors in the (10^{-8}- 10^{-12} s^{-1}) time-scale and thus enable to confirm qualitatively the backbone rigidity and flexibility. A sequence similar to the one described in (139) was used, enabling to use the ratio of peaks intensities of a proton saturated experiment over a reference experiment as a measure of the NOE effect. A 5 s recycling delay was used.

miR-155 precursors secondary structure analysis The Laboratory frame Nuclear Overhauser Enhancement Spectroscopy SpectroscopY (NOESY) is a 2D experiment used to study the NOE effects. In NOESY experiments, this effect leads to the presence of cross peaks at the intersection of the frequencies of the spin that interacts by cross relaxation. The NOE effect between two spins at a distance r is proportional to $\frac{1}{r^6}$ which usually restrains observable interactions for spin spaced by less than 6 A. H,H-NOESY is commonly used to assign the imino protons involved in base paring of nucleic acid, enabling to study the base pairing patterns. H,H-NOESY experiment has been used for the assignment of RNA imino protons, enabling to get information of the secondary structure of the miRNA precursors.

12.3 Study of some chemical exchange processes by NMR

12.3.1 Chemical exchange regimes in NMR

Chemical exchange of a set of spins is a process in which the spin undergoes at least two different chemical environments as a function of time. The different chemical environments create at least two states for the spin considered. Chemical exchange can be induced by a transient interaction with exchange of free and bound state (inter or intra molecular) or by interconversion of the system with exchange between different conformations (intramolecular). The effects of chemical exchange on NMR spectra depend on the properties of the interactions and on the NMR acquisition parameters. A short introduction of typical exchange regime in NMR is given here. The common case of exchange between two states is introduced in this paragraph. Let's consider one spin and let A and B the two states in exchange. The two states are characterized by their chemical shift in rad.s^{-1} (ω_A and ω_B) in one dimension and by the exchange rate constant k_{ex}, sum of the average first order rate of exchange from state A to B and B to A (in s^{-1}).

The effects of the exchange of the NMR spectrum depends mainly on the frequency difference $\Delta\Omega = |\omega_A - \omega_B|$ which itself depend on the spectrometer field strength B_0, the exchange rate k_{ex}, the nature of the nuclei and also the populations of each states. The differences of magnitude between k_{ex} and $\Delta\Omega$ determines three classes of exchange regime :

- slow exchange. This is a limit situation for $k_{ex} \ll \Delta\Omega$. In this situation, there is no significant exchange during the time required for NMR signal detection. Thus the signals of the two states are detected at their respective chemical shifts. Providing the NMR experiment enable each state to fully relax, linewidth and populations correspond to the one of each state. EXSY (Exchange SpectroscopY) experiments enable to detect which peaks are in exchange.
- fast exchange. This is a limit situation, for $k_{ex} \gg \Delta\Omega$. In this situation, there is rapid A/B interconversion leading to averaging of the peak parameters during the detection period of the NMR experiment. The averaged signal is characterized by population-weighted chemical shift and relaxation rates.
- intermediate exchange corresponds to all the situations between fast and slow exchange limits, including the situation $k_{ex} = \Delta\Omega$ characterized by coalescence of the two peaks, which set the limit for the slow and fast intermediate exchange. The closer to coalescence, the broader the peaks are and this can lead to a complete disappearance of any observable signal.

12.3.2 Chemical shift perturbation for the study of TRBP/miR-155 precursors interactions

An important use of secondary NMR observables (chemical shift, magnitude and peak linewidth) in biology is for the study of the interactions between two macromolecules. Let consider we look at the spectrum of one entity alone and then at the same spectrum with another molecular entity added in the NMR tube. If there is an interaction there will be a change of the chemical environment of some nuclei, that can lead to a change of one or more of the 3 secondary NMR observables of peaks detected in the first spectrum. A large application of this strategy has been done in the present work using chemical shifts. The variations of ^{1}H and ^{15}N chemical shifts in ^{1}H-^{15}N-HSQC like experiments is a common way to quantify the magnitude of the binding per protein residue, and was used in this work. These chemical shifts constitute two probes per residue backbone but their variances are different. A solution to combine both variables in a single indicator has been proposed and called the compound chemical shift (202). The idea is to scale one of the two chemical shifts with the average standard deviation of the two chemical shifts (σ_{HN} and σ_H) estimated from the experimental Biological Magnetic Resonance Bank (BMRB) data. Then the formula of the Euclidean distance was used to create a monodimensionnal shift, giving the following formula :

$$\Delta\delta_{comp} = \sqrt{\Delta\delta_{HN}^2 + \frac{\Delta\delta_N^2}{\frac{\sigma_{HN}}{\sigma_N}}} \text{ with } \frac{\sigma_{HN}}{\sigma_N} = 6.5 \tag{12.1}$$

12.3.3 NMR Exchange Spectroscopy Experiments (EXSY)

12.3.3.1 Principle of EXSY experiment

NMR is an effective method to monitor the rate of interconversion of systems that are in slow equilibrium on the NMR time scale. An asymmetric exchange between two states called F and U will be considered here. The corresponding equilibrium is noted :

$$\mathrm{F} \underset{k_{UF}}{\overset{k_{FU}}{\rightleftharpoons}} \mathrm{U}$$

where k_{FU} and k_{UF} are respectively the forward and reverse average first order kinetics constants (in s^{-1}). These constants correspond to the frequency of the transitions and not their velocity. When two spins in a molecule are in slow exchange on the NMR time scale with k_{ex} in the range 0.2 s^{-1} to 10^2 s^{-1}, with each state detectable by NMR, an experiment called Exchange Spectroscopy (EXSY) or ZZ-exchange spectroscopy enable to study the exchange rates, populations and relaxation rates between the exchanging states.

Let's consider two sites $\mathrm{S_F}$ and S_{U} corresponding to nucleus of spin $\frac{1}{2}$ with state F and U with possible polarization states α and β, characterized by their chemical shift in rad.s^{-1}

ω_F and ω_U and by their fractional populations P_U and P_F ($P_\mathrm{U} + P_\mathrm{F} = 1$). As homonuclear NOESY, EXSY is based on the transfer of polarization between two spins. In the absence of scalar coupling between the two spins F and U two mechanisms can lead to this transfer of polarization, chemical exchange and cross relaxation :

- Chemical exchange changes the site of the spins ($\mathrm{S_F}$ to S_U ; $\mathrm{S_U}$ to S_F).
- Cross relaxation corresponds to the result of a process in which the molecular motion of a spin at one site interacting via dipole-dipole interaction with other spins can induce polarization transition at other sites. This leads to the experimental NOE effect. In HN-EXSY, the transfer occur on the nitrogen for which the polarization transfer due to cross relaxation effects are negligible compared to the transfer associated with the chemical exchange process ($\mathcal{A}_{EXSY}1$) because of the low gyromagnetic ratio of ^{15}N, 10 times lower than ^{10}H.

Sequence description of the EXSY experiments, with the classical equations used to analyze EXSY data are given in (184; 75), with a density matrix description in (184) and detailed examples of application are given in (76) and (146). Two reviews on diverse aspects of EXSY experiments can be found in (185) and (219). The NMR review (144) presents EXSY experiments among a broader range of NMR methods used to study dynamics by NMR. The main features of the EXSY sequence are briefly summarized here **fig. 12.2**. Heteronuclear HN-EXSY is based on the transfer of nitrogen polarization between states in slow chemical exchange during a mixing time t_m. First, the magnetization on the amide proton is transferred to the bound nitrogen with an INEPT transfer. z is a dimension parallel to B_0 and ($\vec{x}$,$\vec{y}$,$\vec{z}$) a direct basis with x the dimension along which the NMR signal is detected. The first pulse around y $(\frac{\pi}{2})_y$ creates transverse magnetization. The next delay t_1 (incremented, with a 180° pulse of ^{1}H to decouple the ^{1}H-^{15}N scalar interaction) and the second $(\frac{\pi}{2})_y$ pulse enable to create frequency labeled longitudinal magnetization by the Larmor frequencies of the spins. This creates non equilibrium ^{15}N magnetization along the z dimension. The mixing time t_m allows polarization transfer between the two states F and U in exchange which evolve z magnetization (so the zz-exchange name). Then magnetization is transferred by INEPT back to the amide proton for detection. Depending of the kinetic of the exchange, some undesirable polarization transfer can occur during the second INEPT transfer. Eventually the magnetization is recorded during the detection time t_2.

If a nitrogen is exchanging between states F and U, characterized by $\omega_\mathrm{F}^\mathrm{N}$, $\omega_\mathrm{F}^\mathrm{H}$ and $\omega_\mathrm{F}^\mathrm{N}$,$\omega_\mathrm{F}^\mathrm{H}$, the heteronuclear EXSY experiment yields a 2D spectrum with up to four signals: 2 diagonal peaks FF and UU (autocorrelation) and two crosspeaks FU and UF that reflect the exchanges (F to U and U to F)**fig. 12.2** B. and C.. Standard NMR assignment methods can be used to assign one or the two sets of peaks related to the spins in slow exchange. The EXSY experiment can be also useful to assign one of the state using assignments of the other one. The analysis of classical EXSY spectra require the quadruplet of peaks to be resolved which is rarely the case for all the peaks in a crowded spectrum. Specific labeling and 3D EXSY are attractive strategies to simplify these spectra. Relaxation dispersion experiments based on the monitoring of transverse relaxation (*vs.* longitudinal relaxation in EXSY) covers a broader range of dynamic time scale than EXSY (3.10^2 μs to 10 ms) (144) and can be used to investigate the possibility of faster exchange processes, invisible in EXSY experiments.

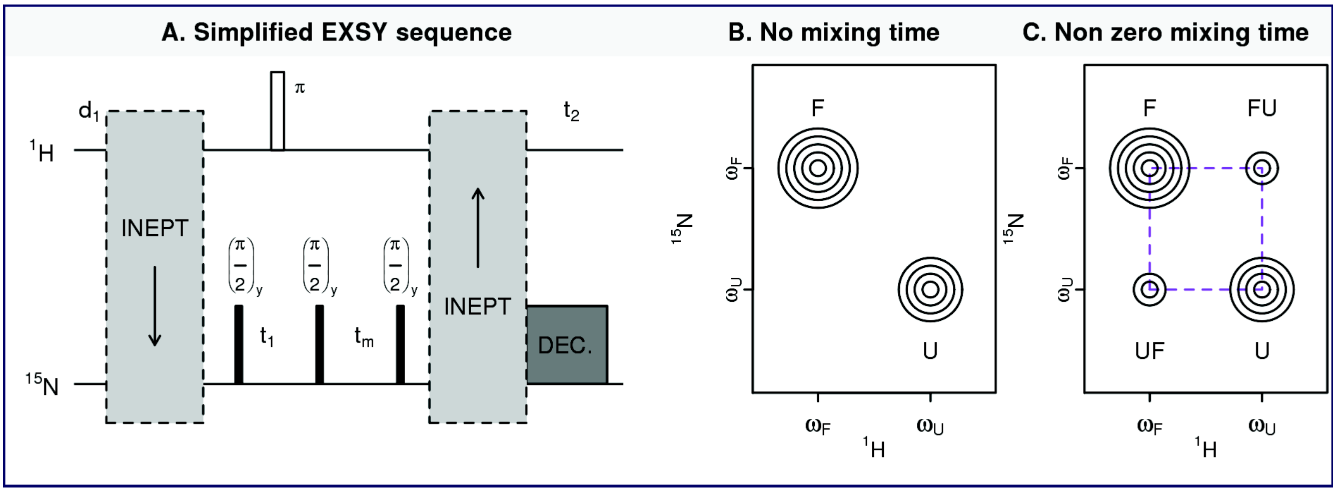


Figure 12.2: **A.** Main steps of an EXSY experiment. The sequence is simplified. INEPT transfers are indicated by gray boxes; water suppression and gradient are not indicated. d_1: recycling delay, t_1 indirect sampling, t_m mixing time and t_2 acquisition time. DEC.: decoupling. **B.** Typical EXSY spectrum with two states in exchange F and U, with no mixing time an assuming no significant exchange during the second INEPT transfer. **C.** Same system than **B.** with a non zero mixing time. Cross peaks FU and UF corresponds to the polarization transfer during the mixing time from F to U and U to F respectively.

12.3.3.2 Exploitation of EXSY experiments

Equations for EXSY data fitting The following notations are used from now (**tab. 12.2**) :

Table 12.2: Notations used in the phenomenological equations describing EXSY peaks intensities

Notation	Description
t	duration of the EXSY mixing time
$I_{\mathrm{XY}(t)}$	volume of the peak corresponding to the transfert of polarization from states X to Y for the mixing time t
$\mathrm{X}(0)$	volume of the peak from states X before the start of the mixing period t
R_1^X	rates of the longitudinal magnetization return to the equilibrium (s^{-1})
R_2^X	rates of the transverse magnetization return to the equilibrium (s^{-1})

Equations to fit EXSY datasets can be conveniently derived using phenomenological equations of nuclear magnetization including chemical exchange, the Bloch-McConnel equations (190). This strategy requires first that the exchange occur between weakly coupled systems (117) ($\mathcal{A}_{EXSY}2$). These equations are theoretically directly applicable to the EXSY data but this would include the fit of bi-exponentials that can be unreliable with experimental dataset (117). EXSY data fitting is usually performed with approximate solution for the following slow exchange condition ($\mathcal{A}_{EXSY}3$) :

$$k_{\mathrm{FU}} + k_{\mathrm{UF}} \ll |i(\omega_U - \omega F) + (R_2^{\mathrm{U}} - R_2^{\mathrm{F}})| \tag{12.2}$$

Under the approximation of fully relaxed states ($\mathcal{A}_{EXSY}4$) at the end of the experiments and without significant exchange during the INEPT transfer ($\mathcal{A}_{EXSY}5$), application of $\mathcal{A}_{EXSY}3$ leads to the following system of equation describing the evolution of the total intensities (volumes) of each peak (formulas from (75)) :

$$\begin{cases} a_{11} & = k_{\mathrm{FU}} + R_1^{\mathrm{F}} \\ a_{22} & = k_{\mathrm{UF}} + R_1^{\mathrm{U}} \\ a_{12} & = -k_{\mathrm{UF}} \\ a_{21} & = -k_{\mathrm{FU}} \\ \forall i \in [\![1,2]\!],\ \lambda_i & = \frac{1}{2}\left[(a_{11}+a_{22}) + (-1)^{3-i}\left[(a_{11}-a_{22})^2) + 4k_{\mathrm{FU}}k_{\mathrm{FU}}\right]^{\frac{1}{2}}\right] \\ I_{\mathrm{FF}}(t) & = I_{\mathrm{F}}(0)\frac{1}{\lambda_1-\lambda_2}\left[(\lambda_1 - a_{11})e^{-\lambda_2 t} - (\lambda_2 - a_{11})e^{-\lambda_1 t}\right] \\ I_{\mathrm{UU}}(t) & = I_{\mathrm{U}}(0)\frac{1}{\lambda_1-\lambda_2}\left[(\lambda_1 - a_{22})e^{-\lambda_2 t} - (\lambda_2 - a_{22})e^{-\lambda_1 t}\right] \\ I_{\mathrm{FU}}(t) & = I_{\mathrm{F}}(0)\frac{1}{\lambda_1-\lambda_2}\left[a_{21}e^{-\lambda_1 t} - a_{21}e^{-\lambda_2 t}\right] \\ I_{\mathrm{UF}}(t) & = I_{\mathrm{U}}(0)\frac{1}{\lambda_1-\lambda_2}\left[a_{12}e^{-\lambda_1 t} - a_{12}e^{-\lambda_2 t}\right] \end{cases} \tag{12.3}$$

EXSY data fitting and implementation EXSY experiments, like NOESY experiments require high sensitivity. Accurate quantification of the populations in EXSY experiment require to have both state in exchange fully relaxed, especially when the relaxation properties of the two states that undergo chemical exchange are different. However if one of the state is weakly populated and since the cross peaks intensities will be below the level of the ones of the minor state. As multiple EXSY experiments with various mixing times are required for an estimation of the kinetics constants a tradeoff has often to be made between the requirements for an accurate quantification and the possibility to measure something. To be able to record enough signal for the cross peaks, high sample concentration were used (around 1 mM) with the BEST (Band-selective Excitation Short-Transient) versions of the EXSY experiment implemented by Bernhard Brutscher (unpublished at the time of writing). Both 2D BEST-EXSY and a 3D BEST-EXSY sensitivity enhanced version were used in this work. Fitting of EXSY data require a series of EXSY experiments at different mixing times including a 0 s mixing time with all other sequence parameters identical.

12.4 Non trivial indiscernible process in chemical exchange study by NMR

Some of the main indiscernible processes in NMR for the study of chemical exchange and relevant for the present study are commented here :

- Presence of chemical exchange can be analyzed with atomic site resolution by the set of NMR observables (chemical shift, peak magnitude, peak linewidth and indirect measurements including relaxation rates, NOE, paramagnetic relaxation enhancement and residual dipolar coupling). However absence of detectable differences in one observable does not imply there is not exchange between multiple states in the system conditions used. For instance the chemical shifts for an observed nucleus undergoing the exchanging state can be equivalent.
- Local conformational exchange and effects of interactions are indistinguishable using chemical shift. A rigid part of a macromolecule can be close to dynamic part which creates a time dependent chemical shift in the rigid region. This will resemble conformational exchange in the rigid part. Chemical shift perturbation for the study of interactions will represent both interacting nuclei and conformational changes associated with the interaction.
- The linewidth of the peaks in NMR are proportional to the size of the macromolecules. So without special design of the experiment, a mixture of a small stable monomeric macromolecule with a large macromolecule or a macromolecular complex (> 100 kDa), the large species might not be detected without adequate/dedicated experiments. The study of large systems (> 60 kDa) by NMR is an active research area using specific experiments and labeling strategies.
- For one spin with more than two states in intermediate or in fast exchange, the fact the peak parameters depends on the weighted properties of the peaks of that would appear in slow exchange often makes the determination of the properties of the different states a challenge; the data can suggest the exchange phenomenon is simpler that what it actually is.

Part III

Production of TRBP and miR-155 constructs

Chapter 13

Production of the RNA binding region of TRBP

13.1 Selection of the production strategy: *E. coli*

Possible ways to produce human proteins include human cells, recombinant strategies in another host (in particular *E. coli*) or chemical synthesis with cell free systems (*in vitro* strategy) using isolated cellular pathways. Various considerations have motivated the use of the production strategy :
- *Yield*. For the biophysical studies aimed to be performed, in particular NMR, large amount of proteins in the order of milligrams are required. Among the three types of methods mentioned, the use of human cells is expected to give much lower yield than the other methods for the same amount of money spent in the protocol. Production in *E. coli* or using cell free systems can give comparable yields for the same cost despite the former system often give much higher yield for human proteins non toxic for *E. coli*.
- *Ease of stable isotope labeling for NMR*. NMR requires for many experiments enrichment of the proteins by stable isotopes (^{15}N, ^{13}C and ^{2}H in this study). The corresponding labeling media are very costly for human cells or for other eukaryotic systems like insect cells. Production in *E. coli* and *in vitro* enable generally much lower labeling costs than eukaryotic systems for comparable yields. In *E. coli*, NH_4Cl and glucose can be used as the sole source of nitrogen an carbon respectively and the cells can grow in $^{2}H_2O$ solvent.
- *Post-translational modifications*. Many human proteins are systematically modified after their translational, including phosphorylations and glycosylations. Production in *E. coli* and standard *in vitro* synthesis do not include these modifications. A study performed on the protein TRBP in HeLa cells (and insect cells) has shown that TRBP exists as a mixture of multiple phospho-isoforms, the most abundant form being however the non-phosphorylated protein (215). There are four phosphorylation sites on serines, two by inter dsRBD regions (215). The results of this study are also in favor of the absence of other large post-translational modifications. Thus, study of the non-phosphorylated form of TRBP appeared relevant. In this respect,

the choice of a system that will not produce an ensemble of variably phosphorylated proteins, that strongly complicate biophysical studies, appeared interesting.

These considerations led to the choice of *E. coli* for the protein production in this work. The strain BL21(DE3) was used since it is widely considered to be an appropriate strain for high-level protein production (104). The choice of *E. coli* as a production system has determined all the purification strategy, set up in order to use the advantages of this system and to alleviate some related problems.

13.2 Expression vectors for the production of TRBP constructs in *E. coli*

A poly-histidine tag (`MAHHHHHHGHHH`) has been use for efficient and quick purification by Immobilized Metal Chelate Chromatography (IMAC, NiNTA here). As already mentioned, tags were considered undesirable for this study as they can alter the properties of the protein sequence of interest. They were systematically removed before the protein analysis by in solution biophysical techniques. A cleavage site (`QLENLYFQ|G`) for the Tobacco Etch Virus (TEV) has been placed just after the poly-histidine tag in order to be able to remove this tag to get the desired TRBP sequence. A glycine residue remains at the N-terminus after cleavage. Another design was used to produce TRBP1-dsRBD1 with a maltose binding protein (MBP) followed by a thrombin cleavage site (`LVPR|GSH`). For this construct only used in a specific part of the thesis, 3 residues (`GSH`) remains after the thrombin cleavage. The cleaved construct design to include TRBP-dsRBD1, TRBP1-dsRBD1 (long isoform), TRBP-dsRBD2 and TRBP-dsRBD1&2 are named respectively TRBP-D1, TRBP1-D1, TRBP-D2 and TRBP-D1D2. The tagged version names are the same with the prefix "polyH-TEV".

The synthetic nucleic acid sequence of the selected regions of TRBP were optimized for expression in *E.coli* taking into account the different codon bias in this species compared to humans (despite this codon bias may play a role in the kinetic of protein folding of some organism).

The resulting sequences were synthesized chemically and inserted into a plasmid PLXR06 with kanamycin resistance or a pMAL vector with ampicillin resistance for TRBP1-dsRBD1 by the RoBioMol platform at the Institute of Structural Biology in Grenoble (IBS).

13.3 Cultures of ^{1}H-^{15}N labeled RNA binding region of TRBP in *E. coli*

1. Transformation of E.coli cells For the transformation of *E. coli* and the first preculture, the common cell culture rich medium Luria Broth (LB) was used. A solution of 20 g.l^{-1} LB media and 15 g.l^{-1} agar was prepared and autoclaved. All the following steps were performed in sterile conditions using either gas flame or a sterile hood. At a temperature of about 50°C, used to solubilize the solution, the LB-Agar media was mixed with a concentrated stock of antibiotic (kanamycin or ampicillin) in order to reach a 30 μg.ml^{-1} concentration. The moderate temperature avoids too much degradation of the antibiotic molecule. The plates were then stored at room temperature until full solidification of the liquid. In parallel, aliquots of 40 μl of competent BL21DE3 cells were mixed with 2 μl of plasmid solution and incubated for 30 min in 1.5 ml sterile closed tubes placed on ice. Then the tubes were heat shocked into a 42°C water bath for 45 s. The tubes were then stored on ice at 4°C for 15 min. Eventually 360 μl of liquid LB media was added to each tube and they were incubated at 37°C during 1 hour. 100 μl samples of the small cell cultures were then dispersed on LB-agar plates, and incubated 16 h at 37°C. Two controls were systematically performed using the same protocol without adding plasmids to the cells or without plating the cells. They were systematically negative.

2. Minimal media preparation In order to obtain labeled proteins, the main cultures were performed in a minimal media where the sources of ^{15}N and ^{12}C or ^{13}C are provided by defined components. The composition of the common minimal media M9 used is given in **tab. 13.1**. The first part of the M9 media was prepared by mixing phosphate salts, sodium chloride and [U-^{15}]-labeled NH_4Cl (Cambridge Isotope Laboratories) in distilled water before an autoclave step. The resulting sterile solution was supplemented with minerals from sterile stock solutions whose concentration are reported in **tab. 13.2**. The latter minerals could not be added before the autoclave step to avoid their precipitation. The media was supplemented with vitamins from a sterile stock solution **tab. 13.3** with 2 ml per liter of culture. The stock solution of vitamins is prepared by adjusting the pH to 7 previously to a filtration at 0.22 μm. The pH of the solution is then reduced to about 5 to get a higher preservability of the solution. The vitamins would have been degraded by the autoclave step if they would have been added before it. Then 4 g.l^{-1} of [U-^{12}C] glucose or 3 g.l^{-1} [U-^{13}C]-labeled glucose (Cambridge Isotope Laboratories) when required, were dissolved in a fraction of the medium, filtered at 0.22 μm and reinjected in the medium. 30 μg.ml^{-1} of antibiotic (kanamycin or ampicillin) was added to the solution before the inoculation step. The initial pH of the medium is about 7.

Table 13.1: Composition of the autoclavable basis of the M9 media used.

Compound	**mass concentration (g.l^{-1})**
Na_2HPO_4, $7H_2O$	10
KH_2PO_4	3
NaCl	0.5
$^{15}NHCl_4$	1

3. Precultures Four to eight isolated colonies on LB-agar plates were used to inoculate a first preculture of LB media of 5 to 20 ml with 30 μg.ml^{-1} antibiotic (kanamycin or ampicillin). The latter culture was incubated at 37°C with agitation at 180 rpm for 6 hours. The LB preculture

Table 13.2: Concentration of the mineral supplemented to the M9 medium.

Sterile solution	Concentration (mM)	Added volume (ml.l_{culture}^{-1})
$MgSO_4$	1000	1
$CaCl_2$	100	1
$MnCl_2$	100	1
$ZnSO_4$	50	1
$FeCl_3$	100	0.5

Table 13.3: Composition of the stock solution of vitamins supplemented to the M9 medium.

Compound	stock mass concentration (g.l^{-1})
Pyridoxin	25
Biotin	25
Panthothenate, hemi calcium	25
Folic acid	25
Choline choride	25
Niacineamide	25
Riboflavine	2.5
Thiamine	125

was used to inoculate a volume v of [U-^{15}N]-labeled or [U-^{15}N,U-^{13}C] M9 minimal media to an initial optical density (OD) at 600 nm, OD_{600}, of 0.05 to 0.1. This second preculture was incubated overnight at 37°C with agitation at 180 rpm and the volume was chosen in order to have enough cells to inoculate the main culture, expecting a final optical density of about 2.5

4. Main cultures and cell storage The desired volume of M9 media for the main culture was inoculated with sufficient M9 pre-culture to give a starting optical density (OD) at 600 nm, OD_{600nm} of 0.05 to 0.1. The cultures were incubated at 37°C with agitation at 180 rpm. The incubation was performed until OD_{600nm} reached 0.7 to 0.8. The cultures were then stored at 25°C for 45 minutes. Protein expression was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM by adding a sterile stock solution of 1 M IPTG in water. The induced cultures were incubated at 15°C for 18 hrs. The incubation at low temperature was used since it is the recommended temperature to favor high yields of soluble proteins according to a specific study (273) and a large scale study performed by multiple structural biology consortium and structural genomic centers (104). This empirically observation is likely to be related to a slower protein production enabling recombinant proteins to fold more properly. Moreover, preliminary experiments have suggested that the yields were better after a long induction at 15°C compared to an induction at 37°C. The final OD_{600nm} was in the range 3.5-4.5. The culture was harvested by centrifugation and the cell pellet stored at -80°C until required.

13.4 Cultures of U-^{2}H labeled TRBP-D1, TRBP-D2 & TRBP-D1D2

13.4.1 Culture of [U-^{2}H,^{13}C,^{15}N]-labeled TRBP-D1D2

Double colony selection A study reported that the recombinant protein production of isolated *E. coli* colonies in ^2H_2O media was heterogeneous (250). A protocol to select the best colonies regarding protein yield was proposed by performing a double colony selection (250). This protocol was adapted for the present study, leading to the following steps :
- Transformation of plasmid in BL21DE3 *E. coli* cells
- Culture in LB-H_2O media at 37°C to OD_{600nm} = 0.7
- Spreading on LB-^2H_2O plates
- Incubation of plates at 37°C during 36h
- Selection of 5 colonies with the largest diameters on the plate and of a smaller one so that the diameter distribution between selected colonies is heterogeneous, which is important to see a selection effect.
- Cultures of the selected colonies in LB-^2H_2O media at 37°C to $OD_{600nm} = 0.7$. 6 colonies were used to inoculate two similar series of 6 cultures.
- Induction overnight at 15°C with agitation
- Estimation of OD_{600nm} of the 6 duplicated cultures and protein production on gel for one series of 6 cultures.
- Spreading on LB-^2H_2O plates of the 6 remaining cultures
- Incubation of plates at 37°C during 36 h
- Cultures from biggest colonies in LB-^2H_2O media at 37°C to $OD_{600nm} \approx 0.7$
- 20 % glycerol stock stored at -80°C for future use, as recommended in (250) to avoid the long selection step before each new culture.

The results of the main steps of the double colony selection for TRBP-D1D2 are given in **fig. 13.1** with details in the legend. The protocol enabled to significantly increase the yield of the protein by a factor of about 2 (2.3) compared to a bad colony choice (colony 6). All the cultures of TRBP-D1D2 in ^2H_2O were performed from the cell stock in glycerol. The cells were grown in M9 media prepared with [99.85 %-^{2}H]-labeled water and 3g.l^{-1} [U-^{2}H,^{13}C] glucose and 1g.l^{-1} $^{15}NH_4Cl$. Applied on TRBP-D1D2 this protocol enabled to get about 300 μL of purified protein at 500 μM per liter of culture (3.4 mg.l^{-1})

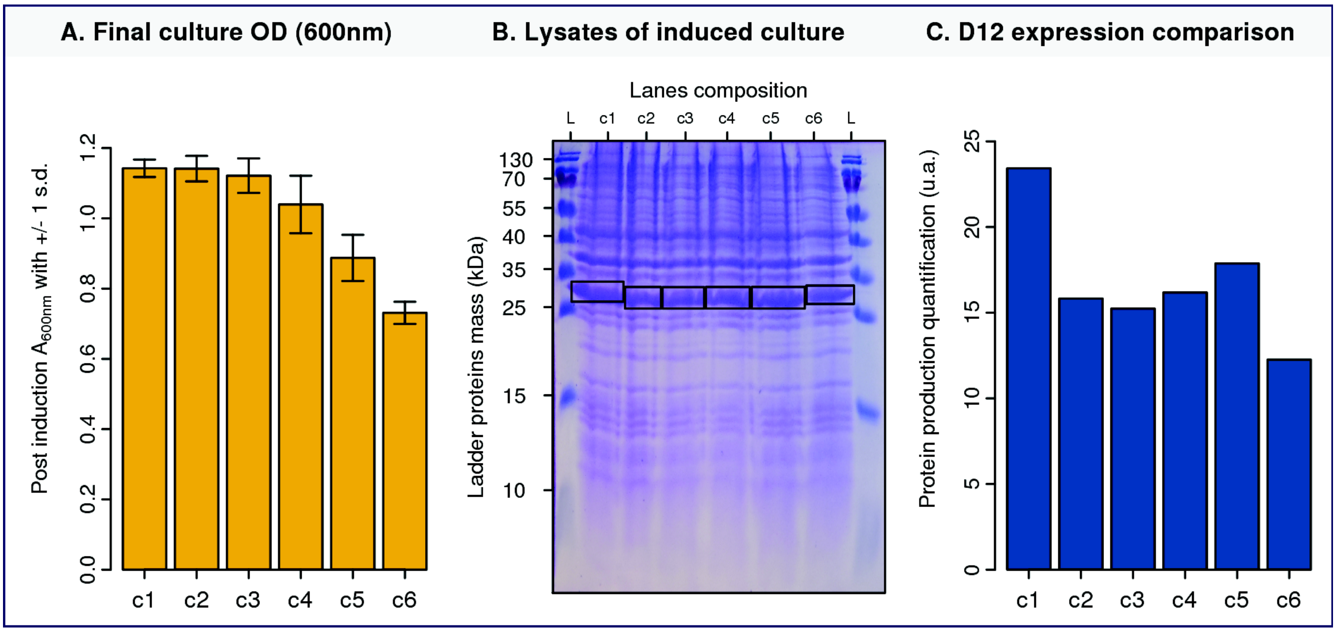


Figure 13.1: First round of colony selection for their growth ability and TRBP-D1D2(19-228) production in *E. coli* BL21DE3. **A.** Comparison of the final OD_{600nm} of the largest colonies selected on LB-D_2O plates and grown in LB-D_2O media. Each initial culture was spitted into 2 sub-cultures and 3 measurement of the final OD were performed for each culture. The error bars represents +/- one standard deviation of these six measurements. **B.** 15 % acrylamide SDS-PAGE gel of lysates of cultures 1 to 6 after induction. Lysis was performed by incubating the cultures in the SDS-PAGE loading buffer and passing them 6 times in a syringe capped with a long narrow needle. The amount of material loaded on the gels was normalized to take into account the different concentration of cells at the end of each cultures. **C.** Quantification of the intensity of the TRBP-D1D2 band in each lane obtained by summing pixel intensities in the boxes of **B.** for the ones over the background intensity.

13.4.2 Cultures of *U*-^{2}H-^{15}N labeled TRBP-D1, TRBP-D2 & TRBP-D1D2

Cultures of [*U*-^{2}H,^{15}N]-labeled TRBP constructs were performed without double colony selection for TRBP-D1 and TRBP-D2. Multiple colonies of TRBP-D1 and TRBP-D2 were used to inoculate pre-cultures of LB media and then M9 media, both prepared in H_2O. The culture were then used to inoculate a M9 media prepared in [50 %-^{2}H]-labeled water to progressively adapt the cells to ^2H_2O. Cells from this culture were then used to inoculate a preculture in M9 media prepared with [99.85 %-^{2}H]-labeled water, 4 g.l^{-1} unlabeled glucose and 1 g.l^{-1} $^{15}NH_4Cl$. The stock of TRBP-D1D2 prepared after double colony selection was used from this step, just after a preculture in LB prepared in [99.85 %-^{2}H]-labeled water. For TRBP-D1 and TRBP-D2, the yields were rather low: for one liter of culture of the single domains of TRBP, about 100 μl at 100 μM for TRBP-D2 (0.08 mg.l^{-1}) and 200 μL at 100 μM for TRBP-D1 (0.1 mg.L^{-1}) were obtained after purification. That's why after the first culture, some deuterated ^{15}N Isogro® has also been used for the single domains at the induction time with a concentration of 0.5 g per liter, in order to get a higher yield (increased by at least 4 times).

Purification comment The deuterated proteins were purified like the non deuterated protein, at room temperature in protonated solvents to allow exchange of amide deuterons to protons. The purification protocols are developed in the following paragraphs.

13.5 Development of a cell lysis strategy

Soluble protein purification Human recombinant proteins expressed in *E. coli* are produced in the cytoplasm of the bacteria and a fraction of the proteins often form inactive aggregates called inclusion bodies. The latter phenomenon usually results from the lack of appropriate folding conditions, folding machinery, processing enzymes, availability of partners and/or very high protein concentration in the bacterial cell. As a result there are usually at least two fractions of protein in the bacterial cytoplasm: soluble proteins and protein in inclusion bodies. The repartition of proteins in the two fractions depends on multiple factors, including the properties of the protein produced, cell expression conditions, bacterial strain, etc. As the protein in inclusion bodies are often misfolded and are hard to dissociate without altering their structure, two general purification strategies are commonly used to purify recombinant proteins in *E. coli*: purify the soluble protein in native conditions or use denaturing or mild denaturing conditions to solubilize the fraction of protein in inclusion bodies. In the latter case, the proteins will be unfolded and will have to be refolded. Protein refolding is usually a long, inefficient and time consuming step. Indeed, the resulting protein fold might not be the same on obtained after the native purification. The protein TRBP is predicted to have flexible regions (**section 5.1.1**) so the refolding strategy would complicate the analysis since the experimental evidence of flexible regions could be a real property of the protein or a mis-refolded region. For these reasons the protein were purified under native condition. In this case, it is hoped but not guaranteed that the fold of the protein will be very similar to the one in human cells; functional tests can help to detect problems.

13.5.1 Criteria for the development of a purification strategy

A protocol had to be set up in order to screen the protein constructs. Various steps are required to produce a protein recombinantly: culture, cell lysis and protein purification. There are at least 4 key requirements when some screening is performed :
- no volume constraint: the method should be applicable for any sample volumes.
- volume independent result: the best conditions found for small scale cultures must also work for large scale ones.
- parallelizability: the method should be easily performed in parallel for multiple cultures or fast enough to apply in order to minimize any incubation time effects between different samples.

- damage free: as for any technique, the method should avoid to alter the fundamental properties of the proteins under study.

13.5.2 Comparison of cell lysis strategies

Although it is easy to scale the culture and purification columns, it is harder to lyse the cells with a scale independent strategy. This can be an issue since cell lysis, a key step in the purification, aims to extract the protein in the *in vivo* experiment to an *in vitro* environment where the protein should be stable. Various cell lysis methods are commonly used for recombinant protein expression in *E. coli*: sonication, high pressure, freezing and thawing cycles, mechanical cell disputers or chemical lysis. The main advantages and limits of the previously cited methods will be briefly compared here to argument the choice of the selected strategy :

- *Sonication*. This method is based on ultrasounds to break the bacterial cell wall. The method is an efficient way to lyse *E. coli* cells and to break nucleic acids. This is an advantage for proteins interacting with nucleic acids when it is desired to get the protein alone. However, sonication has various drawbacks. It is usually difficult to reproduce the same effect with various sonicators. The main drawback is that the proteins can be damaged and form aggregates. A study of the effects of sonication on various unrelated proteins with different sequences has been performed (255). The latter work also report many other works leading to similar side effects consequences of sonication. To avoid such a risk that could impact both the protein properties and the evaluation of the protein concentrations, sonication was not used in this biophysical study (condition "damage free" potentially not matched).
- *Freezing and thawing cycles*. Repeated cycles of freezing and thawing can contribute to cell lysis. This is not a very efficient cell lysis method (about 50% efficiency with a standard protocol, (131)) and the repeated formation of crystals during each cycle, responsible for the lysis, can damage the proteins. For these reasons, this method was not selected here (condition "damage free" not matched).
- *Mechanical, bead-vortexing and pressure-based methods*. These methods include the widely used French press, Eaton presses and Microfluidizers. These methods are efficient but are time consuming, not parallelized in the standard design and not applicable for small scale samples like 50 ml cultures (conditions "no volume constraint" and "parallelizable" not matched).
- *Chemical lysis*. Various components can be added to a cell pellet to trigger lysis:

  - *Enzymes*, like lysozyme can disrupt the cell wall of Gram negative bacteria. The mechanism *in vitro* mainly involves disruption of some small part of the cell wall that diminish locally the resistance of the wall to osmotic pressure. The cell will eventually be broken a release its content into the solution.

  - *Detergent*. They disrupt lipids and protein interactions that are required for the cell stability. Some small ionic size detergents like SDS denature proteins and are thus not of interest to release the soluble proteins whereas many non ionic detergents have been used to solubilize proteins and for cell lysis. Non ionic detergents have also often the advantage to increase the solubilization of proteins but can also solubilize aggregates. A chemical lysis using enzymes and non denaturing detergents fulfills the 4 conditions mentioned before. It is easy to adjust the volume of lysis buffer to the amount of cells to be lysed so the conditions "no volume constraint" and "volume independent result" work. The parallelizability just requires to mix pairs of solutions. And the damages are limited by the use of gentle detergents. A key requirement would be to be able to remove the detergents which might be difficult for proteins with large hydrophobic surfaces. Eventually because of the mechanism of action of the detergents the chemical lysis method might not be the preferred one if the goal is to purify complexes formed in the bacteria. As the current project is not concerned by the last two limits and because of the fulfillment of the 4 conditions, this method has been preferred.

13.5.3 Setup of a chemical lysis buffer for *E. coli* BL21 DE3

Setup of the lysis buffer recipe One widely used commercial reference for a detergent-based *E. coli* cell lysis is BugBuster®. It is a lysis buffer based on detergents with a patented composition using octyl-glucoside detergents. Together with its cost (12.8 €.$od_{600nm}^{-1}.l^{-1}$ without nucleases), these properties motivated me to use a home made lysis buffer. The set up of this buffer was not performed with a large screening of many compounds but mainly by merging the knowledge of various compounds and buffers expecting additivity effects. The build up of the lysis buffer recipe is detailed step by step here :

- *buffer basis*. Previous experience in yeast proteomics had suggested me that the buffer Hepes at pH 7.5 with 100 mM KCl and 10 % glycerol worked well to solubilize many proteins. The pH should be adjusted to match solubility conditions of the recombinant proteins produced, which was the case here (pI[TRBP-D1D2] = 9.5). Hepes has a low pH temperature dependence which allows to incubate at different temperatures (4 to 30°C) without significant pH effects. Glycerol helps to reduce protein/protein interaction and increases protein solubility.
- *Lysozyme* from chicken egg white was used at 1 $mg.ml^{-1}$ concentration to contribute to the lysis. In order to facilitate lysozyme cleavage the lysis was performed at 30°C.
- *Triton X-100* is a commonly used non-ionic detergent (0.6 kDa) that was used at 0.1 % (v/v). It is compatible with the activity of many enzymes including lysozyme and nucleases. It is often used at low concentration for *in vitro* synthesis of macromolecules using isolated cellular machinaries.
- *DNAse I and RNase A*. The detergent doesn't cleave the DNA and this can be an important issue since the DNA makes the lysate solution highly viscous. The viscosity can impede further

column based purification steps. Thus some DNAse was added to the lysis buffer at 25 μg.ml^{-1}. Since TRBP is known to bind RNA, to avoid co-purification of protein and RNA some RNase at 25 μg.ml^{-1} was used to cleave RNA (single stranded).

- *Calcium and magnesium chloride*. DNAse I and RNase A require respectively calcium and magnesium for the reaction they catalyse. Both were used at 10 mM.

- *Antiproteases*. In order to limit the effects of *E. coli* proteases on the recombinant proteins, pepstatin and leupeptin were used at 1 μg.ml^{-1}.

- *Imidazole* at 10 mM was added to the lysis buffer in order to have a molecule that can compete with the weakly bound protein on the NiNTA resin. This enables to get higher yield in the first purification binding step. The addition of imidazole can inhibit some lysozymes (259) but the presence of this molecule is considered very important for the purification step and was added in the lysis buffer for convenience.

Protocol for cell lysis

(1) Preparation of the lysis buffer. The lysis buffer was made just before use but could also be stored at -20°C and used after thawing in a room temperature water bath. The latter method appeared to work well, was very convenient and mainly used for tests.

(2) Cell pellets resuspension in lysis buffer (50 mM HEPES pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 10 mM Imidazole, 10 % (v/v) glycerol, 1 mg/ml lysozyme, 0.5 % triton X-100, 1 μg/ml leupeptin, 1 μg/ml pepstatin, 20 μg/ml DNase and 20 μg/ml RNase).

(3) In order to standardize the amount of lysis buffer to the amount of cells, the volume of lysis buffer $V_{\text{lysis buffer}}$ used was standardized to the measured optical density at 600 nm at the end of the culture $OD_{600\text{ final}}$ of volume V_{culture} using the following ratio :

$$v_{\text{lysis buffer}} = \frac{v_{\text{lysis buffer}} * V_{\text{culture}}}{40} \tag{13.1}$$

(4) Cell lysis was performed at 30° C for 40 minutes with agitation (180 rpm) in a closed glass bottle of volume about 5 times the volume of lysis buffer in order to ensure an efficient agitation.

(5) 5 mM β-mercaptothanol was then added to avoid protein oxidation. This reducing agent couldn't be added before because it would have reduced the key disulfites bound that structure lysozymes.

(6) The lysis reaction was clarified by centrifugation at 30000 $\times g$, during 30 minutes at 4°C.

Effects of Triton X-100 and glycerol concentration on the lysis In order to understand the potential effects of glycerol, Triton X-100 and lysozyme on the lysis efficiency, a series of lysis has been performed on a non-induced protonated *E. coli* culture with the standard recipe or without triton or with fraction of the glycerol used **fig. 13.2**.

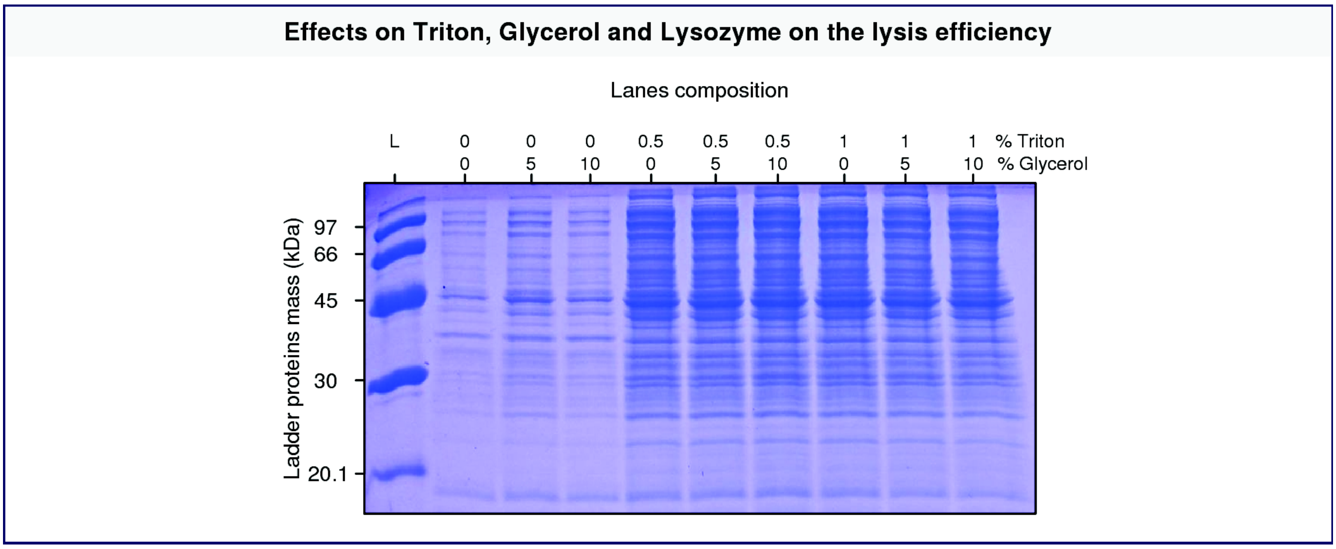


Figure 13.2: Effects of Triton X-100 and glycerol concentration on the lysis of a non induced culture of *E. coli* BL21DE3 analyzed by SDS-PAGE. The percentage (v/v) of both Triton X-100 and glycerol were varied for different lysis.

The results show that the lysozyme has a moderate effect on the lysis, the key component is Triton X-100. The efficiency of the lysis by Triton X-100 does not depend on the glycerol concentration in the range 0% (v/v) to 10% (v/v). Doubling the concentration of Triton-X100 does not lead to an increase of the lysis efficiency in the range tested. Both glycerol (10% (v/v)) and lysozyme (0.5 % (v/v)) were kept in the recipe for their expected roles, protein solubilization and degradation of bacterial cell walls respectively.

Comparison of the lysis recipe with other lysis techniques The commercially available BugBuster® was a good reference with which a home made lysis buffer could be compared with. Since sonication is an efficient lysis method compatible with small samples, it was also used as reference for cell lysis. 20 ml Cultures of protonated TRBP-D1D2 (19-228), TRBP-D1 (19-99) and TRBP-D2 (157-228) were performed as previously described in **section 13.3** and were complemented with 10 g.l^{-1} unlabeled Isogro® in order to facilitate the observation of the recombinant proteins in lysates by SDS-PAGE. To have a comparison of broader significance for dsRBD containing proteins, the results for TRBP are complemented by similar construct of PACT produced with the same method with a poly-histidine with a TEV site N-terminal fusion in a PLXR06 plasmid: PACT[32-102], PACT[124-313], PACT[32-196] and PACT [239-313] as well as TRBP[257-366]. The cell pellets of each culture were separated into 3 equal parts and resuspended in 1.5 ml lysis buffer. They were lysed with 3 different methods: BugBuster®, home made lysis buffer and sonication. To avoid general buffer effects, the basis of the lysis buffer for each lysis method was the home made lysis buffer including antiprotease inhibitors and nucleases but without glycerol and Triton X-100. To avoid a differential effect coming from the presence of lysozyme in the home made buffer, this enzyme was added at the same concentration in the three lysis buffers. Method specific protocol features and additional components are listed for each method (called A, B and C) :

- A: 1× BugBuster® (using a 10× stock), with a one hour incubation at room temperature (≈ 25°C) and regular homogeneization by reversing the tubes.
- B: home made lysis buffer with one hour incubation at 30°C with agitation at 180 rpm.
- C: Sonication in home made lysis buffer with lysozyme, nucleases and glycerol but without Triton-X100. Four times 10 seconds impulsion pulses were alternated with 20 seconds breaks on ice.

After lysis, the samples were centrifuged during 30 min at 30000 $\times g$ at 4°C. The pellet was resuspended in water and homogenized with multiple passes through a narrow needle using a syringe. The supernatant and resuspended pellets were then prepared for SDS-PAGE as previous described and loaded on gels. The Lysozymes (14.3 kDa) and the acrylamide percentage of the gel (15 %) do not enable to clearly resolve the single domains constructs in a few cases. The results are given in **fig. 13.3**.

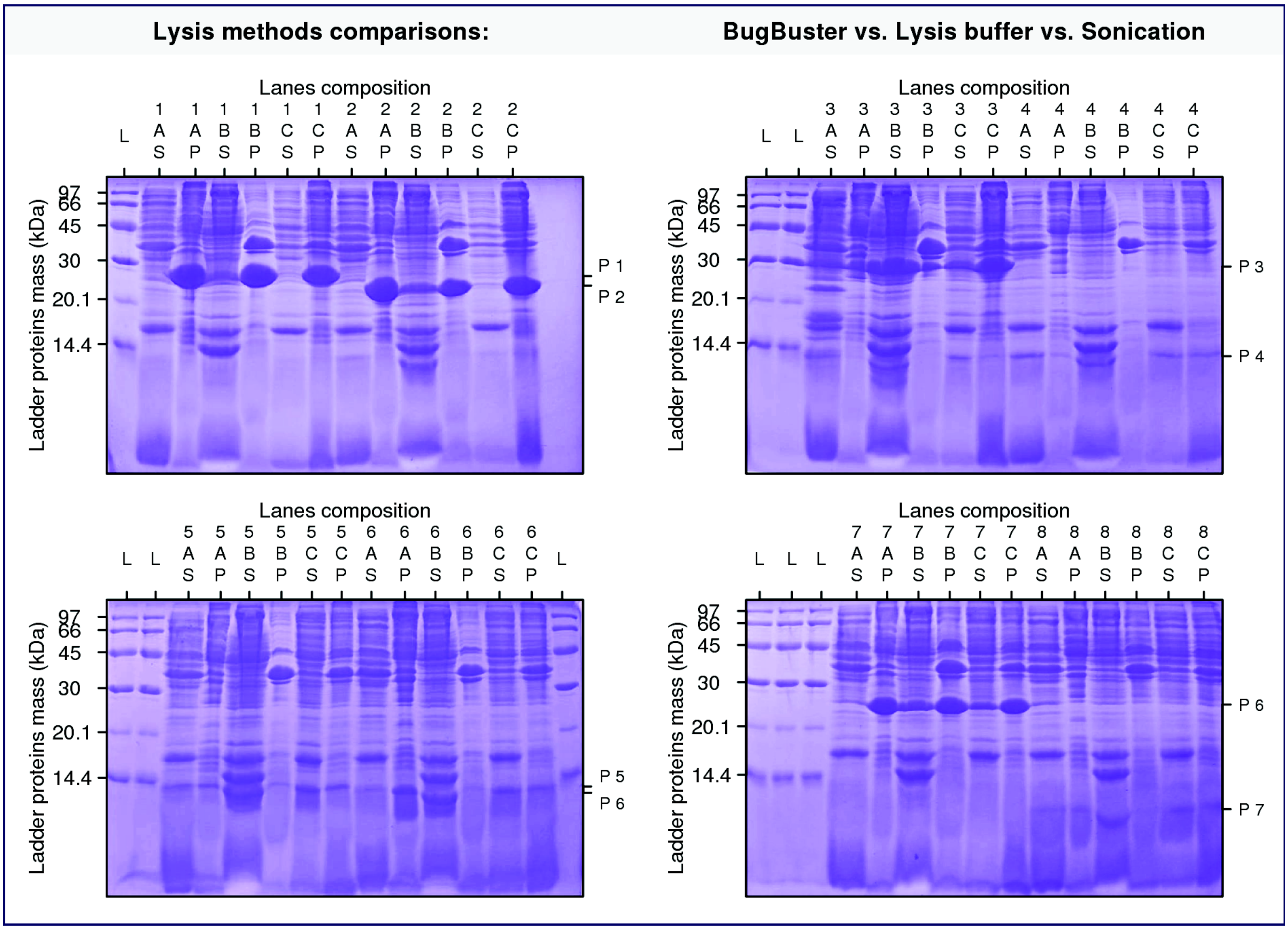


Figure 13.3: Comparison by SDS-PAGE of the efficiency of three lysis methods (A: BugBuster®, B: home made lysis buffer, C: sonication) on different protein constructs of TRBP and PACT. The methods are performed as described in the main text. Ladder lanes are described with a "L". The composition of the other lanes are described by three letter, indicating from top to bottom the protein construct (1 to 9), the lysis method used (A, B, C) and the fraction loaded (S: Soluble, P: from resuspended pellet). The vertical positions of the recombinant constructs of TRBP or PACT are indicated on the right axis. [Polyhistidine-TEV site]-tagged proteins codes: 1: TRBP[257-366], 2: PACT[32-196], 3: TRBP[19-228] (TRBP-D1D2), 4: TRBP[19-99] (TRBP-D1), 5: TRBP[157-228] (TRBP-D2), 6: PACT[32-102], 7: PACT[124-313], 8: PACT[239-313])

The gels on **fig. 13.3** show that both chemical lysis buffers are more efficient than the sonication conditions used despite a large proportion of the proteins is not solubilized with all the techniques, probably because of their production in inclusion bodies. The protocol of home made lysis buffer appears even more efficient than the BugBuster® used in these conditions and for these proteins. This might not only be linked to the composition differences between the lysis buffers but also to the shaking (180 rpm) and temperature (30°C) that were used with the home made lysis buffer and not with BugBuster®. Because of the low cost (*cf. infra*) of the home made lysis buffer, its clear and controllable composition and its efficiency for these proteins, it was used for all the cell lysis in the work reported here.

Cost comparison with BugBuster® The main costly components in the buffer used are the nucleases, leading to a cost of 1.5 €.od$_{600nm}^{-1}$.l^{-1} with the previous protocol. Using the recommended volume of 1 × BugBuster® (Novagen), 5 ml for 1 g of wet cell pellets and assuming that 1 l at OD$_{600}$ $= 1$ corresponds to a wet pellet weight of 1.7 g, with 330 euros for 50 ml 10× product this imply a cost of 12.8 €.od$_{600nm}^{-1}$.l^{-1}. With the same amount of nucleases, the BugBuster strategy cost about 10 times (14.2 €) the cost of the home made buffer. And this factor is a minimum as it seems that more BugBuster is useful to better lyse the cells. In the Novagen protocol, they mention a BugBuster/Culture ratio for small cultures of 1/20, which corresponds to a cost of at least 75 €.OD$_{600nm}^{-1}$.l^{-1} assuming a culture OD$_{600nm}$ over 1.

13.6 NiNTA purification

The supernatant were passed over 4 ml of Ni-NTA affinity chromatography resin (QiaGEN) per culture liter. The column was pre-equilibrated with storage buffer (50 mM HEPES pH 7.5, 100 mM KCl, 10 mM MgCl$_2$, 10 mM CaCl$_2$, 10 mM Imidazole, 10% (v/v) glycerol and 5 mM β-mercaptoethanol). The resin was washed sequentially with 20 column volumes (CV) of storage buffer with 50 mM Imidazole, 20 CV of storage buffer with 2 M NaCl to remove nucleic acids and 20 CV of storage buffer alone. Bound proteins were eluted with fractions of storage buffer with 250 mM Imidazole, starting with a 1 CV fraction followed by pairs of 1 minute incubation and $\frac{1}{4}$ CV fraction of eluted buffer. A typical SDS-PAGE used to monitor each step of the purification is given in **fig. 13.4** for polyH-TEV-TRBP-D1D2 (19-228).

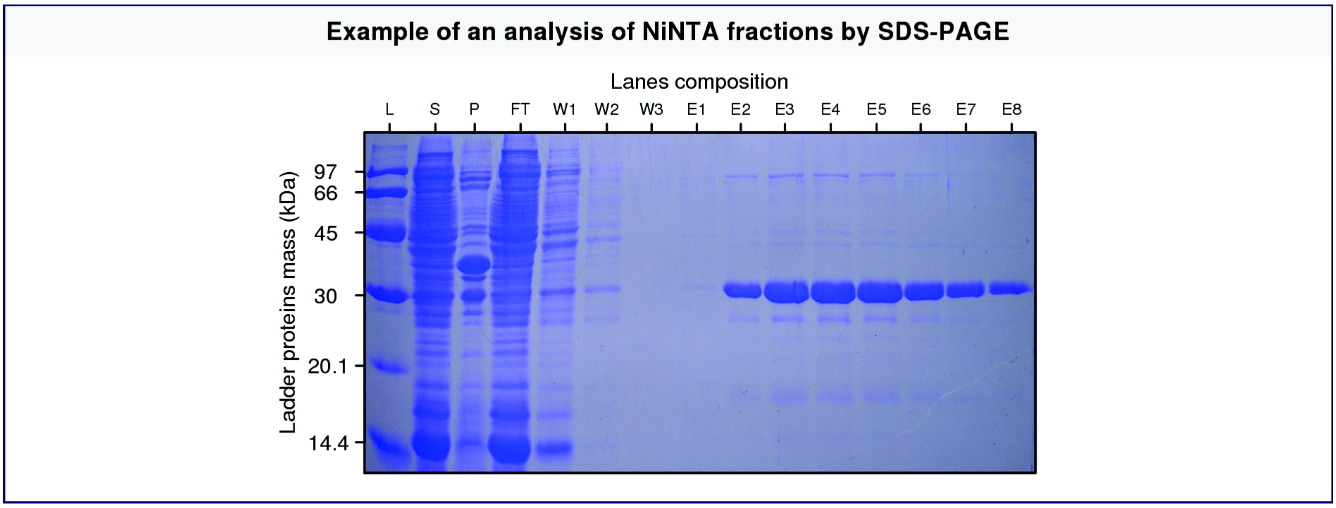


Figure 13.4: Example of the monitoring of sample purity during a NiNTA purifications, corresponding to TRBP-D1D2 (19-228). The purification was performed as described in the main text. L: ladder, S: soluble fraction after home made lysis, P: resuspension of the lyaste pellet obtained after centrifugation of the lysate at 30 000 ×g, FT: column flow through of the lysate, W1: column wash with 50 mM imidazole, W2: column wash with 2 M NaCl, W3: column wash with storage buffer alone, E1-E8 elutions with 250 mM imidazole of 1 column volume.

The ultraviolet spectra of each elution fraction were systematically checked to select the more concentrated fraction and to avoid taking fractions contaminated with nucleic acids. The last fractions were systematically characterized by a high absorbance at 260 nm and were not pooled with the other ones.

13.7 Cleavage of the poly histidine tag

Protocol The observation that the proteins had a tendency to precipitate during the dialysis performed to remove the imidazole after NiNTA and that the cleavage of the histidine tag was inefficient, motivated to perform both dialysis and TEV cleavage at the same time. The eluted protein were incubated with 0.1 equivalents of TEV protease (stored in 50 mM Tris-HCl pH 8.0, 25 mM NaCl, 0.5 mM EDTA, 2 mM DDT and 20 % glycerol). The reaction mixture was dialyzed against 2 × 2 L of storage buffer at 4°C (12 hrs then 4 hrs). The protein solution after cleavage was centrifuged at 3000 $\times g$ during 5 min at 4°C in order to remove some precipitates formed during dialysis. In order to remove the TEV protease, the dialysate was passed over Ni-NTA affinity chromatography resin pre-equilibrated with storage buffer. The same volume of resin than the one used for the first NiNTA purification step was used. The resin was then washed with a series of $\frac{1}{4}$ CV fractions of storage buffer added with 1 min interval. The goal of this step was to bind on the resin the histidine tagged TEV protease, uncleaved proteins and potential contaminants that could have been co-purified during the first NiNTA step.

Cleaved protein purification monitoring by SDS-PAGE An example of gel used to set up the purification conditions of the TEV cleavage with a small case test (1 ml NiNTA) is given in **fig. 13.5**. Additional salt washes on the column with the storage buffer supplemented by 300 mM NaCl followed by an imidazole elution released TEV protein and show that a significant fraction of the proteins is uncleaved. Some cleaved proteins are visible in after the imdazole elution. This might be related with direct or indirect tagged proteins/untagged proteins interactions and/or could be related to protein/column interaction and/or to a cleavage during the elution where both TEV and uncleaved proteins are eluted. Very low amounts of uncleaved protein are detected in the first elutions. From these tests, sufficient quantities of NiNTA (equal to the volume used during the first NiNTA step) and slow loading of the column by the proteins incubated with TEV enable to have only cleaved proteins in the elutions, revealed by the final purification gels **fig. 15.1**.

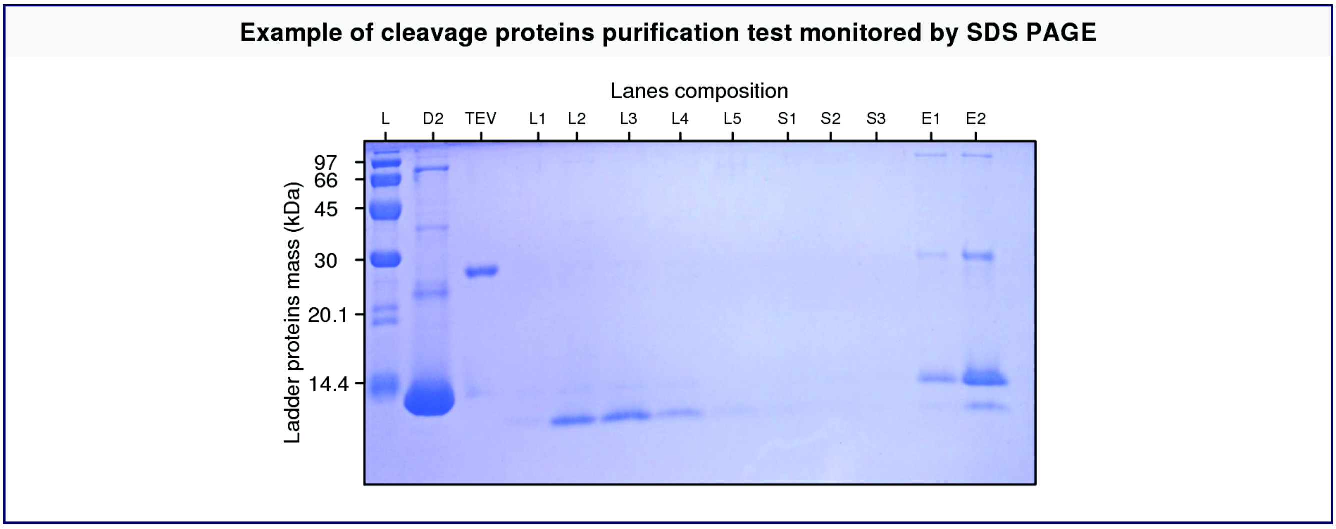


Figure 13.5: Example of a gel used to monitor TEV cleavage, for TRBP-D2. L: ladder. D2: TRBP-D2 sample after NiNTA. One column volume is used for each elution. TEV: TEV protease. L1 to L5: elutions with buffer [50 mM HEPES pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 10 mM Imidazole, 10 % (v/v) glycerol and 5 mM β-mercaptoethanol]. S1 to S3: wash with buffer supplemented by 300 mM NaCl. E1 and E2: elutions with buffer supplemented by 250 mM imidazole. The TEV:TRBP-D2 ratio is this example is 0.05.

It was observed that without tag cleavage the proteins after the NiNTA step had a tendency to aggregate and more than 50% of the protein samples were lost when attempting to concentrate them. This effect was not investigated into detail as the goal was to remove the tag but it shows that in this case the tag removal step is an important step to get sufficient proteins in the end. The effect could be linked to a small RNA contamination present after the NiNTA step that could make the proteins moderately stable at this step of the purification and/or could be linked to the presence of the affinity tag. A tag effect could be related directly to the TEV site and/or to indirect interactions between proteins via the tag like the coordination of ions via poly-histidines. A study on mouse proteins expressed in *E. coli* (151), reports decreased solubility related to the introduction of a TEV site after a polyhistidine tag. The results of this article are more consistent with a causal effect of the TEV site than of the polyhistidines, for their systems.

13.8 Size exclusion chromatography

In order to improve the NiNTA purification steps and to have a first characterization of the samples, a Size Exclusion Chromatography (SEC) step as described in **chapter 7** was added at the end of the purification. This step also enabled to change the storage buffer to a more favorable buffer to study interactions with biophysical techniques. The SEC was performed with no glycerol, to remove it from the media and in phosphate buffer. Phosphate buffer has many advantages :

- invisibity in proton NMR, which is an advantage when the spectrum of small compound is to be measured; it is however highly charged which is a disadvantage for NMR cryprobes. A balance has to be made between this limitation and the need for salts and buffer to have a reasonable environment condition for the interactions.
- three pKa values, enabling to buffer the pH in the range 4 to 8
- very low enthalpy of protonation (ΔH_{ion}[Phosphoric acid, pH = 7.2] = 3.3 kJ.mol^{-1}) (41)) which is a strong advantage to perform Isothermal Titration Calorimetry
- high buffering capacity
- low dependence of pKa on the temperature: at 100 mM the second pKa (7.2) follows $\frac{dpKa}{dT} = -0028$
- no absorption in the ultraviolet range

Size Exclusion Chromatography 23.5 mM of phosphate buffer was used, keeping the 100 mM KCl concentration and the 10mM amount of $MgCl_2$ since it is known that some RNA, especially the large ones, require some $MgCl_2$ to their fold. Since it was not reasonable to adjust the ratio of $MgCl_2$ to the concentration of RNA, the 10 mM concentration was chosen to be 10 time the highest usable RNA concentration. Fractions containing protein were concentrated and 500 μL aliquots were passed over an analytical Superdex S75 10/300 GL column (GE Healthcare) pre-equilibrated with a gel filtration buffer (23.5 mM potassium phosphate pH 6.5, 100 mM KCl, 10 mM $MgCl_2$ and 5 mM β-mercaptoethanol). The chromatograms were characterized by a single well resolved peak. A full analysis of the SEC results is given in **section 15.2.1**.

Concentration The protein-containing fractions were pooled and concentrated with centrifugal filter units to the desired concentration for biophysical studies. 5 mM tris(2-carboxyethy)phosphine (TCEP) at pH 6.5 was added to most protein samples as TCEP has the advantage to be a strong and stable reducing agent.

Chapter 14

Production of the oncogenic microRNA 155 precursors

miR-155 precursors were produced by *in vitro* transcription by Lionel Imbert at the IBS cell free platform. *In vitro* transcription is one of the most common methods to produce RNA and was well developed in the host lab. Recombinant production of RNA in cells using fusion with transfer RNA has also been reported (309). *in vitro* transcription enabled to produce efficiently the amounts of unlabeled RNA required for this study.

14.1 Design of miR-155 precursors constructs for *in vitro* production

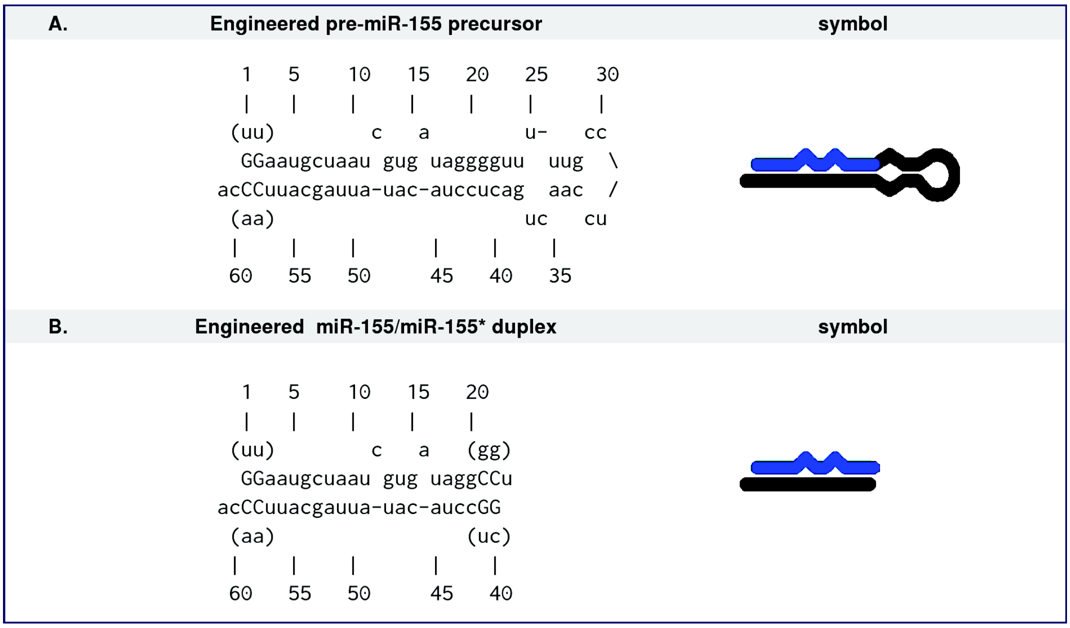
As each translated RNA sequence was required to start with a GG dinucleotide it was necessary to modify the native RNA sequences. Additional sequence changes were required to preserve the Watson-Crick base pairing. The non-native nucleotides are emphasized in the sequences provided in **fig. 14.1**. The related RNA oligonucleotides are: Pre-miR-155 (5′-`ggaau gcuaa ucgug auagg gguuu uugcc uccaa cugac uccua cauau uagca uuccc a`-3′) or miR-155/miR-155* (miR-155: 5′-`ggaau gcuaa ucgug auagg ccu`-3′; miR-155*: 5′-`ggccu acaua uuagc auucc ca`-3′). These sequences were used to design fully complementary double-stranded DNA oligonucleotides containing a T7 promoter (Eurogentec). This are referred to as template oligonucleotides

14.2 Large-scale production of pre-miR-155 and miR-155/miR-155*

- Nucleotide preparation. Frozen template oligonucleotides were incubated at about 22°C during 15 minutes. They were homogenized by pipetting and shortly centrifuged.
- Annealing. Oligonucleotides were then diluted in RNase free water and a 3× concentrated buffer (120 mM Tris-HCl pH 8.1, 3 mM spermidine, 15 mM dithiothreitol, 0.3% Triton X-100, 240 mg.mL^{-1} polyethylene 8000) to reach a 100 μl mix composed of 40 mM Tris-HCl pH 8.1, 5 mM dithiothreitol, 0.1% Triton X-100, 80 mg.ml^{-1} polyethylene 8000 and a DNA concentration to reach 35 μg.ml^{-1} in the final transcription volume of next step. The mix is vortexed to homogenize it and quickly centrifuged. The mix was heated 5 min at 95°C and then incubated 20 min at

about 22°C in order to anneal (hybridization) the complementary parts of the oligonucleotides.

- Transcription mix preparation During the annealing step, a standard T7 RNA polymerase buffer of 5 to 20 ml was prepared: 40 mM Tris-HCl pH 8.1, 1 mM spermidine, 5 mM dithiothreitol, 0.1% Triton X-100, 80 mg.ml^{-1} polyethylene 8000, 4 mM of each NTP and 15-20 mM MgCl^2. The solution was vortex and quickly centrifuged to homogenize the components. Then the proteins were added: 40 μg.ml^{-1} bovine serum albumin, 1 U.ml^{-1} pyrophosphatase and extemporaneously 0.1 mg.ml^{-1} of in-house produced T7 RNA polymerase stored at -20°C.



```
  1   5    10   15   20   25    30
  |   |    |    |    |    |     |
 (uu)        c   a        u-   cc
  GGaaugcuaau gug uagggguu  uug  \
acCCuuacgauua-uac-auccucag  aac  /
 (aa)                     uc   cu
 |    |    |      |    |    |
 60   55   50     45   40   35
```

```
  1   5    10   15   20
  |   |    |    |    |
 (uu)        c   a   (gg)
  GGaaugcuaau gug uaggCCu
acCCuuacgauua-uac-auccGG
 (aa)                (uc)
 |    |    |      |    |
 60   55   50     45   40
```

Figure 14.1: mfold (309) predicted secondary structure of the two engineered microRNA precursors of miR-155, corresponding to the step before and after Dicer cleavage. G and C in uppercase indicate a modified base, used for transcription efficiency. The neighboring base, above or below in parenthesis indicate the native ones. Two symbols representing these RNA constructs are given in the column on the right.

- Transcription The hybridized DNA oligonucleotides were introduced in the transcription mix (final concentration: 35 μg.ml^{-1} DNA matrix) and homogenized by reversing slowly the container tube. The mix was split in 1 ml fractions placed in 1.5 ml Eppendorf tubes. The reaction mixture was incubated at 37°C for 3 hours.

14.3 Purification and folding of RNAs

RNA products were purified by polyacrylamine gel electrophoresis, dialyzed multiple times against water and lyophilized. The folding of each RNA was achieved by heating each sample in water at 96°C for 5 minutes followed by rapid cooling by the addition of ice-cold 2 × phosphate buffer.

Part IV

Characterization of the apo RNA binding region of TRBP and pre-miR-155

The main goal of the thesis reported here is to analyze *in vitro* the interaction of the RNA binding region of TRBP and microRNA precursors, to address the role of each dsRBD in this region, the potential effects of the interdomain region, and whether TRBP can differentiate microRNA precursors before and after their cleavage by the enzyme Dicer. The aforementioned goals require a thorough characterization of the isolated partners involved in the interaction via multiple biophysical techniques. This is the object of this part. The following chapters present the apo characterization of the RNA binding region of TRBP and of the last two precursors of the oncogenic human microRNA 155. This includes an assessment of the homogeneity of the macromolecules involved, some analysis in solution of their potential homo-oligomerization and studies of their general structure and flexibility.

This part contains the following chapters :
- **Chapter 15**: Characterization of the apo RNA binding region of TRBP
- **Chapter 16**: Characterization of apo pre-miR-155 and miR155/miR155*

Structure of the experimental work presentation From here in the chapters devoted to the experimental analyzes, the following organization is adopted, without always an explicit structure :
- *Motivations*, relating the biological question with the selected biophysical techniques.
- *Experimental design*, indicating the samples and special conditions, with references to the biophysical part II.
- *Results*.
- *Data analysis*, including possible explanations of the results using a search of a set of biases and situations that can lead to a similar result. Information from the "Non-trivial indiscernible processes" presented the part II will be used. This part leads to theories that can enhance the knowledge on the system and that can be tested with other experiments.
- *Integration and (partial) conclusions* using a restriction of the set of hypothesis using previously known results and with the remaining potential non checked/non checkable assumption required to analyze the data.

Chapter 15

Characterization of the apo RNA binding region of TRBP

15.1 Homogeneity analysis of recombinant TRBP constructs

Motivations For an appropriate interpretation of the results of the biophysical techniques and thus for the related biological conclusions, an assessment of the purity and homogeneity of the TRBP constructs is required. In addition absence of large aggregates that could indicate irrelevant precipitated form of the protein and absence of nucleic acid contamination are potential biases to be checked

Experimental design and results The purity of the samples was assessed by the combined use of SDS PAGE, SEC chromatography **fig. 15.1** and svAUC for TRBP-D1D2**fig. 15.4**. These methods are complementary: SEC UV signal is easily quantifiable but the resolution is limited and presence of very large aggregates can be undetected on the column used. SDS-PAGE (15 % acrylamide here) allows a better resolution in a defined range of masses and svAUC has a lower resolution but enables to study the whole sample without risk of aggregates filtration. In order to detect potential RNA contamination, an UV spectrophotometer was also used to estimate the ratio between the absorbances at 260 nm over 280 nm of the samples (TRBP-D1D2: 0.47, TRBP-D1: 0.56, TRBP-D2: 0.70). SDS-PAGE and SEC results are given below while svAUC of TRBP-D1D2 is given later in **section 15.2.3**.

Data analysis The SDS-PAGE profiles of TRBP-D1 and TRBP-D2 are compatible with pure proteins. There is a minor low molecular weight species detected in this gel (not reproducible) for TRBP-D1D2 which very likely comes from a cleavage of the protein related to the SDS-PAGE sample preparation as this minor species is not detected in SEC or svAUC. SEC chromatograms are characterized by single peaks far from the exclusion volume (8 ml) and permeation (> 20 ml) volumes of the column used (Superdex S75 10/300 GL column, GE Healthcare) and without other species resolved. Despite similar masses and behavior in SDS-PAGE, TRBP-D1 and TRBP-D2 have a different elution volume, a property which will be commented later. The $\frac{A_{260\text{nm}}}{A_{280\text{nm}}}$ ratios of TRBP-D1D2 (0.47), TRBP-D1 (0.56) and TRBP-D2 (0.70) are compatible with no nucleic acid contamination. A further purification step with a heparin column was used to test whether the ratio could be altered. This purification step is useful when there is nucleic acid contamination in a sample. The resulting ratio did not differ from the one of measured before the heparin step. This is also consistent with no nucleic acid contamination so a heparin step has not been included in the end of the purification protocol. It can also underlined here that the other biophysical analyzes performed in this part and in the next one led also to the conclusion of an absence of nucleic acid contamination for the TRBP constructs.

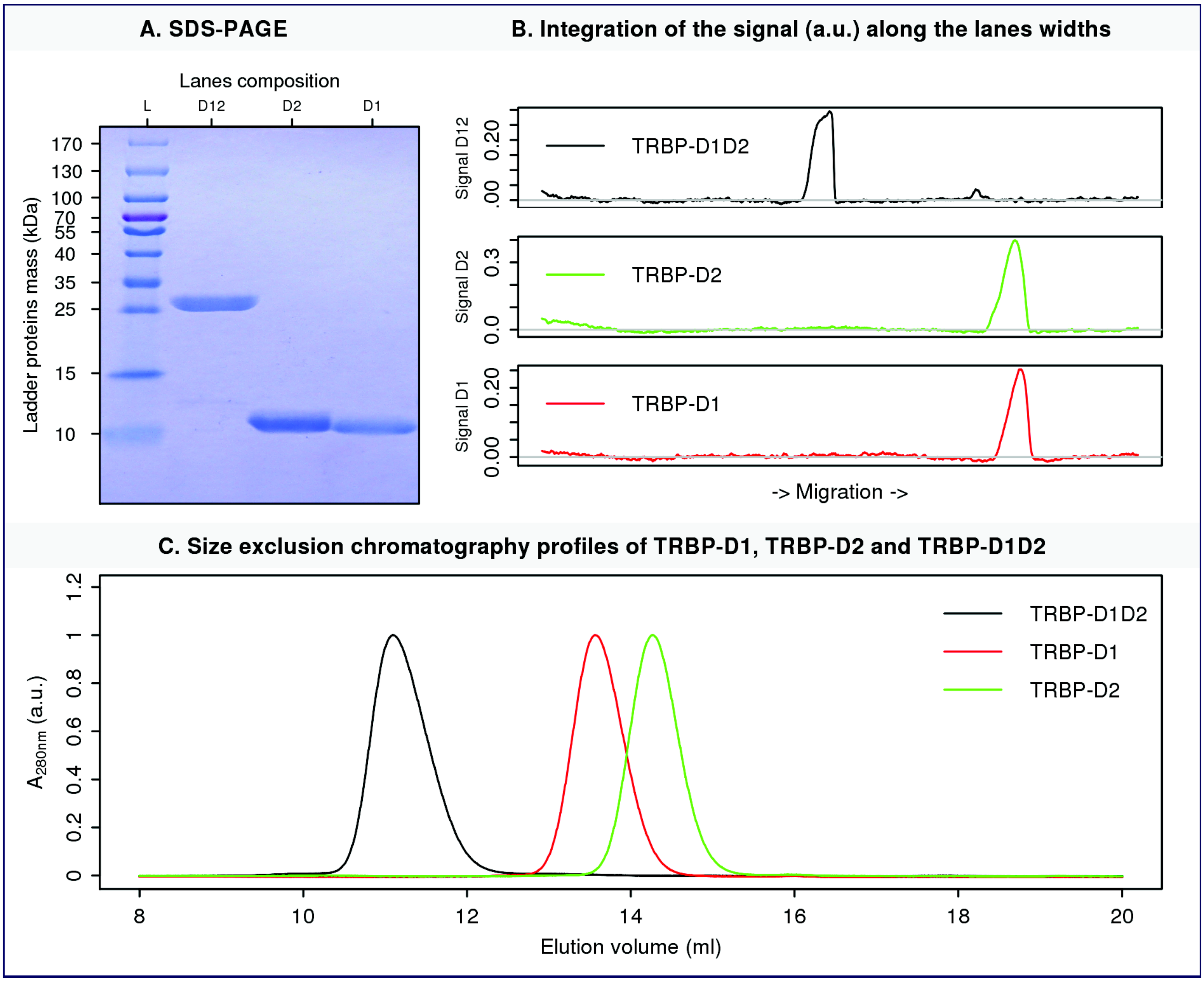


Figure 15.1: Purity analysis of TRBP constructs by SDS-PAGE and SEC at ≈ 25°C. **A.** 15% acrylamide SDS-PAGE gels of purified TRBP-D1D2, TRBP-D2 and TRBP-D1 constructs. **B.** Integration of the signal on the width of the bands. **C.** SEC chromatograms at 280 nm of 500 μl of TRBP-D1D2 (34 μM), TRBP-D1 (140 μM) and TRBP-D2 (43 μM), with normalized intensities.

Conclusion Thus the purity of TRBP constructs was considered to be sufficiently high to apply biophysical techniques to characterize the properties of the RNA binding region of TRBP and its interactions with miRNA 155 precursors.

15.2 Oligomerization analysis of the RNA binding region of TRBP

Motivations The goal of this study is to determine the potential intermolecular interactions occurring for the three type of TRBP constructs produced. The assessment of the oligomerization state of this construct is particularly important for all the subsequent biophysical analysis, including the study of the interaction between TRBP and miR-155 precursors. Three classes of experiments have been successively used to characterize the oligomerization properties of TRBP-D1D2, TRBP-D1 and TRBP-D2 constructs.

15.2.1 Oligomerization analysis with SEC

Serial dilution through SEC experiment Samples of TRBP-D1, TRBP-D2 and TRBP-D1D2 were produced using the standard protocol. A first SEC was performed as described in **chapter 7** as a final purification step, with the samples in [50 mM HEPES pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 10 mM Imidazole, 10 % (v/v) glycerol and 5 mM β-mercaptoethanol]. For each construct, 500 μl of highest concentrated part of the eluate obtained after the SEC part of the purification step was reloaded on the column 4 times in the phosphate buffer used for biophysical studies at $\approx$ 25°C. This corresponded to 34 μM of TRBP-D1D2, 140 μM of TRBP-D1 and 43 μM of TRBP-D2. Since the elution volumes are defined with a random noise error, the four dilutions intend to minimize strongly the risk of graduation of elution volume obtained by chance. Each time, this protocol led to a dilution factor of the protein loaded on the column of about 2.

Results The results of the serial dilution through SEC are presented in **fig. 15.2**.

Data analysis At a 1 ml.min^{-1} flow rate, five points per seconds were recorded for each chromatogram so the elution volume for each peak can be very well estimated by the elution volume for which we have the maximum absorbance of the peak. This leads to the elution volumes stored in **tab. 15.1**.

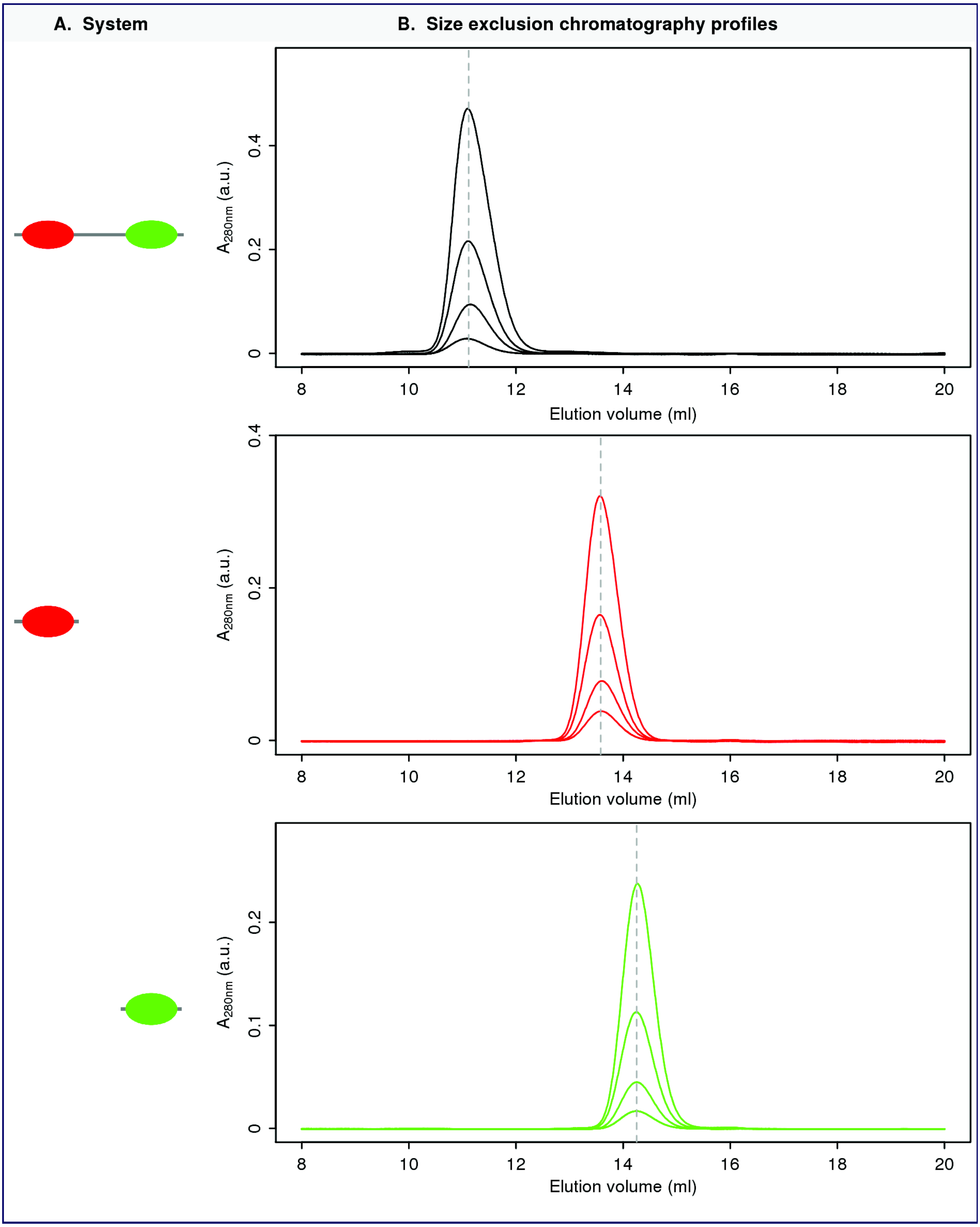


Figure 15.2: Effect of the concentration of TRBP constructs on their size exclusion chromatography elution at ≈ 25°C. **A.** Symbols of the different protein construct used : dsRBD1 in red and dsRBD2 in green **B.** Overlay of the series of 4 size exclusion chromatography profiles at 280 nm. The gray dashed line indicates the position of the mean elution volume averaged over the four experiments.

Table 15.1: Measurement of the elution volume of TRBP constructs for a series of four successive elution-reinjection SEC experiments.

Construction	Experiment	Elution volume (ml)
TRBP-D1D2	1	11.09
TRBP-D1D2	2	11.11
TRBP-D1D2	3	11.15
TRBP-D1D2	4	11.15
TRBP-D1	1	13.57
TRBP-D1	2	13.56
TRBP-D1	3	13.6
TRBP-D1	4	13.67
TRBP-D2	1	14.26
TRBP-D2	2	14.25
TRBP-D2	3	14.25
TRBP-D2	4	14.25

The elution volumes are far from the exclusion (about 8 ml) and permeation limits (over 20 ml) of the column used (Superdex S75 10/300 GL column, GE Healthcare). So the concentration series comparison can be interpreted with the SEC separation principles. If there would be any significant populations of different oligomeric states in equilibrium or more generally significant attractive interactions, we should see a concentration dependence of the peaks or multiple peaks. The non trivial indiscernible processes underlined in **chapter 7** all apply in this situation. The table shows that in the sampled concentration range the elution volumes of the protein constructs are extremely similar. A very slight graduation is observed for the two construct with the first domain of TRBP. However, the variation is extremely small compared to the large difference of concentration used in the experiment. The elution volume comparison does not suggest any significant intermolecular interaction between the domains. Thus SDS-PAGE and SEC strongly suggest that there is essentially one species in these peaks. So providing no bias linked to the SEC matrix, two hypotheses can explain the result observed for each construct :

- the construct is monomeric in the range of concentration sampled
- the construct is a stable oligomer in the range of concentration sampled

For each construct, an easy way to distinguish between these two hypotheses is to estimate the molar mass of the species in solution. This can be achieved with the SEC-MALLS technique.

15.2.2 Oligomerization analysis by SEC-MALLS

SEC-MALLS experiment To estimate the molar mass of the species eluting after the SEC step, a SEC-MALLS was performed for the three TRBP(-D1,-D2,-D1D2) samples at one concentration, in the range of what is use for NMR. 55 μl samples of TRBP-D1 (1.8 mg/mL, 420 μM), TRBP-D2 (1.7 mg/mL, 380 μM) and TRBP-D1D2 (2 mg/mL, 1.6 mM) were loaded on the SEC column. The protocol applied here for the SEC coupled to Multiple Angle Light Scattering in the phosphate buffer at $\approx$ 25°C is described **chapter 8**.

SEC-MALLS data analysis Using the classical data analysis protocol used in MALLS and its multiple assumptions (**chapter 8**), the proposed estimated masses with their standard deviation error are given in **tab. 15.2**.

Table 15.2: Estimation of the mass of the species TRBP-D1D2, TRBP-D1 and TRBP-D2 by SEC-MALLS

Construct	Measured mass (kDa)	Std. deviation (kDa)	Theoritical mass (kDa)
TRBP-D1	8.5	0.5	8.9
TRBP-D2	8.4	0.5	8.2
TRBP-D1D2	23.2	0.9	22.6

To validate this estimation, this result and the previous knowledge on the sample has to be compared with the assumption used to derive the molar mass values :

- $\mathcal{A}_{MALLS}1$. The estimated masses, very close to the monomeric masses, correspond to peptidic chains that are well below the laser wavelenght (690 nm) so this assumption is fulfilled.
- $\mathcal{A}_{MALLS}2, 3$. The assumptions 2 and 3 are hardly testable and their potential effects have been discussed in the methodological part.
- $\mathcal{A}_{MALLS}4$. The concentration independency profile in SEC is compatible with the absence of transient interaction in the species responsible for the detected peaks. This is required for the system to get close to the assumption 4.

So since 2 assumptions might not be perfectly fulfilled, moderate biases are expected. We see however that the weighted mass is very close to the mass of the monomers and non ambiguously far from higher oligomerization species. As a result it seems very safe to conclude that TRBP-D1, TRBP-D2 and TRBP-D1D2 behave as monomers in solution in the concentration range used for the SEC-MALLS study, which encompass the ones of all the other techniques. To check this conclusion using completely different theories and assumptions, an independent study on TRBP-D1D2 by svAUC is presented in the next paragraph.

15.2.3 Oligomerization analysis of the RNA binding region of TRBP by svAUC

svAUC experiment In order to confirm the previous conclusions, a smaller experiment has been performed by sedimentation velocity ultracentrifugation (svAUC). Although less precise than MALLS, svAUC uses a different theories (thermodynamics and hydrodynamics) to estimate the molecular weight of proteins in solution and is thus a good candidate to confirm the monomeric behavior of the RNA binding region of TRBP. The three dimensional NMR experiments have shown that the samples were stable for longer period than the one required for this svAUC experiment (about 15 hours, the first 9 hours being useful for data analysis of this system). The svAUC analysis was performed on two samples of 10 μM and 20 μM using the protocol presented in the methodological part. As usual, the experiment is performed in the same phosphate buffer than in the others biophysical studies and at 25°C

svAUC results The concentration profiles recorded during the experiment at 280 nm are given in **fig. 15.3**.

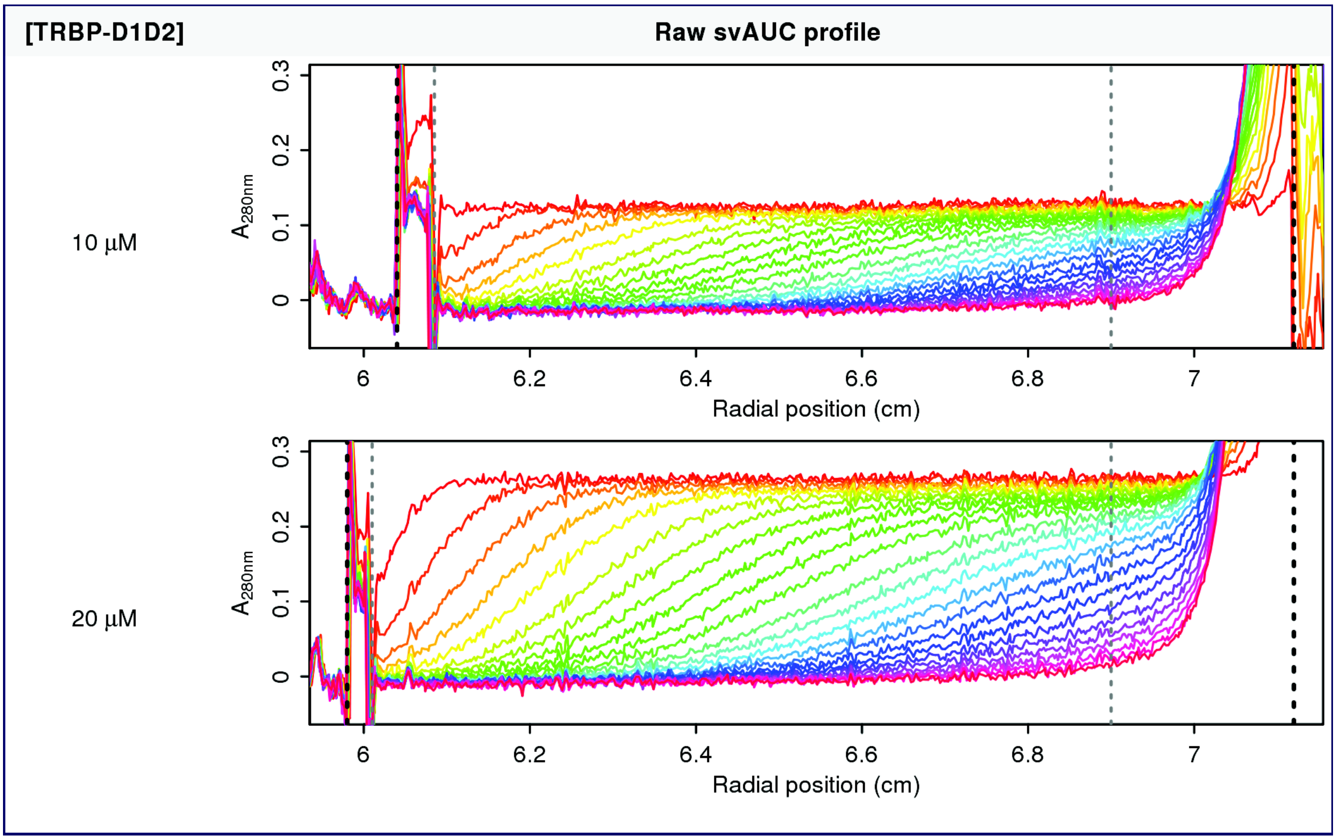


Figure 15.3: Superimposition of the 25 first scans recorded at 280nm in the svAUC experiments for TRBP-D1D2 at 10 μM and 20 μM. The bold dashed lines indicate from left to right the position of the liquid meniscus and the position of the cell bottom. The two lighter lines in gray indicate the zone used to fit the data. The right light gray dashed line indicate an approximate position after which the effects of the cell bottom became obvious in the sedimentation profiles

svAUC data analysis The result of the diffusion deconvoluted $c(s)$ model fitting is given in **fig. 15.4**. The $c(s)$ profile indicates that the experimental sedimentation coefficient s_{exp} is concentration independent in the modest low concentration range tested and its value is 2.19 S. Although only two concentrations are used, this "concentration independency" is perfectly consistent with the one shown more rigorously in a broader range by SEC.

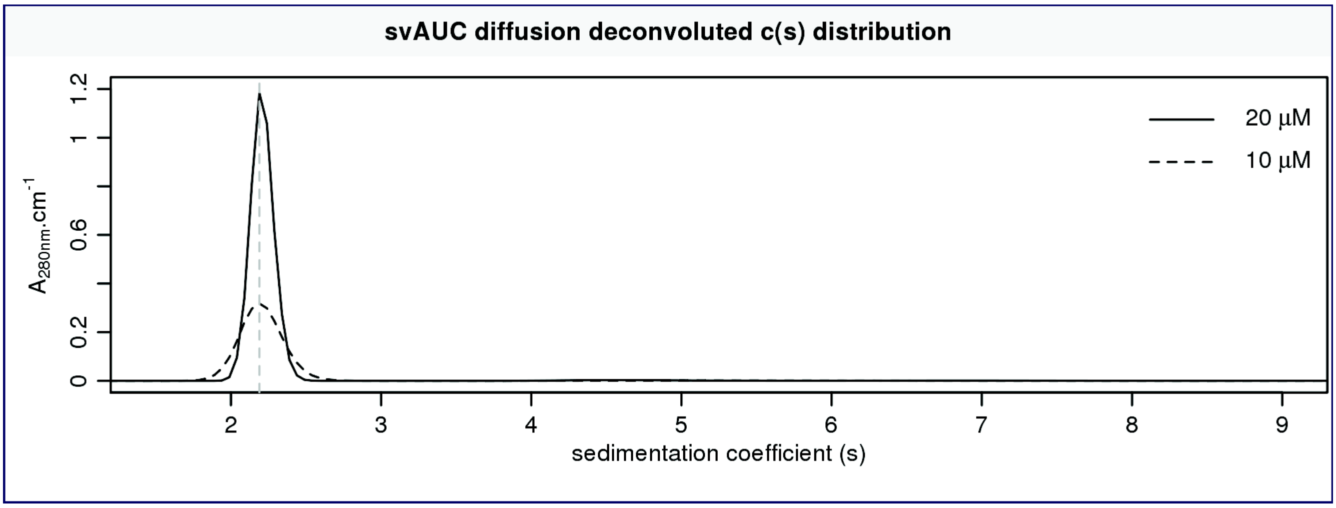


Figure 15.4: Superimposition of the diffusion deconvoluted $c(s)$ profiles of TRBP-D1D2 at 10 μM and 20 μM respectively. The dashed gray line is positioned at the mean of the maximum of both distributions.

As a result of this concentration independency of s_{exp} of TRBP-D1D2 i.e. no intermolecular interactions, a modeling of the sedimentation with the Lamm equation with a single non interacting sedimenting species can be applied. This model, without assumptions about the shape of the sedimenting species, leads (via the classical Svedberg equation) to a molar mass estimation of 22.4 kDa which is close to the real mass of 22.6 kDa. Taking into account the assumptions used in the derivation of the model used (**chapter 11**) this estimation is considered very good. As a result, this svAUC analysis confirms the monomeric behavior of the RNA binding region of TRBP. Additionally, fixing the mass in the c(s) model to fit the data leads to a frictional ratio of 1.43. This value is larger than the one expected for a spherical particle (1.25) and means the protein has an elongated shape in solution. This property will be commented later in the light of the NMR results. This frictional ratio value is in excellent agreement with one previously reported in the literature (1.43) for the region of human PKR containing its two dsRBDs linked by the 23 residue flexible linker (270).

15.2.4 Validation of the relevance of the oligomerization study by 2D-NMR

A previous study in the literature has shown that folded structures of the first and the second dsRBD of TRBP could be obtained by X-ray crystallography and NMR respectively (290). If the proteins used here are fully unfolded, it would suggest an inconsistency between the protein produced in this study and the results obtained in the literature. Thus the monomeric property of the protein produced has a general meaning providing their also have folded regions in solution. The moderate size of the monomeric constructs below 211 residues allows to study qualitatively their fold using standard BEST-TROSY experiments. A spectrum for each construct is given in **fig. 15.5**. The high dispension for most peaks suggest constructs mainly folded.

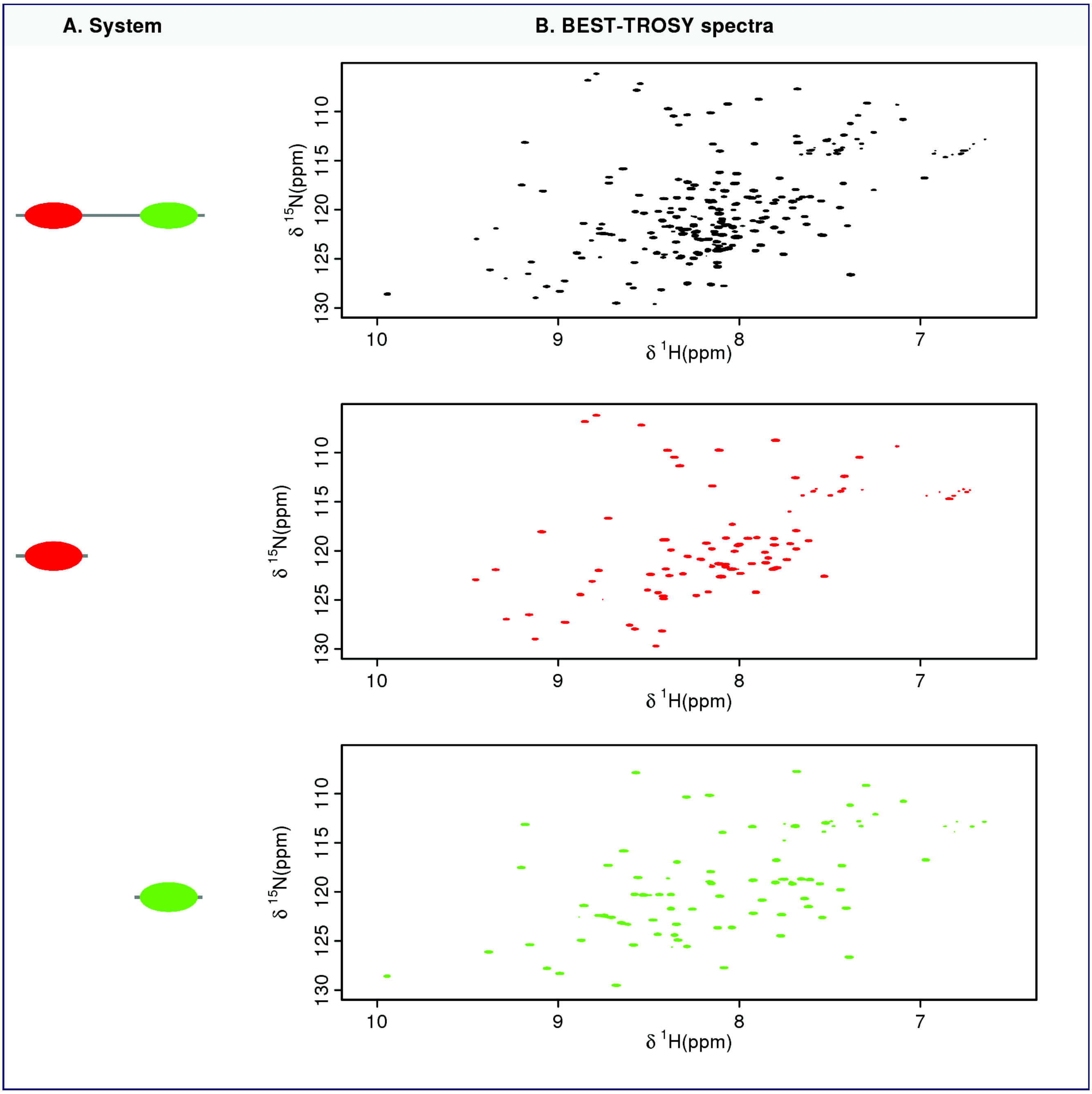


Figure 15.5: BEST-TROSY spectra of TRBP constructs recorded at 25°C at 600 MHz. The constructs are symbolized by cartoons : dsRBD1 in red and dsRBD2 in green.

Conclusion Thus the SEC experiments have shown that the RNA binding region of TRBP and its two isolated dsRBDs in highly purified samples forms stable single excluded species in the physico-chemical conditions used. Quantification of the mass of these species with two independent methods, MALLS and svAUC (for TRBP-D1D2), show unambiguously that they are monomeric. Additional NMR experiments show that this property is not linked to an obvious potential bias corresponding to a production of fully unfolded constructs that would be incompatible with previous data on TRBP.

15.3 NMR study of the secondary structure and dynamics of the RNA binding region of TRBP

NMR was used to characterize at atomic resolution in solution the backbone of the RNA binding region of TRBP and to compare it the presence of miR-155 precursors. This should enable to reveal properties of the TRBP / miR-155 precursors interactions(s). NMR experiments are the key methods in this study to analyze the behavior of each dsRBD and of the interdomain region in the apo form and with miR-155 precursors. To reach this goal, a relation between peaks in NMR spectra and pairs of atoms has to be performed and is called NMR spectrum assignment. This chapter presents the NMR backbone assignment of the RNA binding region of TRBP required to analyze interaction by NMR, and the high frequency dynamics of the backbone HN vectors. The latter information can be used to determine the folded regions of the proteins and the (CA, CB, CO) chemical shifts complement this information by giving the secondary structure propensity of the residues. Apart from clearly mentioned cases, the results presented in this part have been recorded in the phosphate buffer [23.5 mM potassium phosphate pH 6.5, 100 mM KCl, 10 mM $MgCl_2$ and 5 mM β-mercaptoethanol, 5mM TCEP].

15.4 NMR Assignment, secondary structure and dynamics of TRBP2-D2

15.4.1 Assignment of TRBP-D2

Experiments and results A set of BEST experiments has been used to assign the [^{1}H,^{15}N] TRBP2-D1 resonances: HN(CO)CACB, HNCACB, HNCO and inter-residual HADAMAC experiments. Figures representing spectra used for this assignment as well as an assigned HSQC spectra is given in **fig. 15.6**. The number of non side chains peaks in the spectrum of TRBP2-D2 constructs (67) matches well the expected number according to the sequence (73). 91.8% of the backbone [^{1}H,^{15}N] resonances have been assigned; the missing residues are the three first ones including the residual glycine. The residues without [^{1}H,^{15}N] resonances assigned are represented in red in the following sequence of the construct and the prolines with no amino groups in gray :

`GECNPVGALQELVVQKGWRLPEYTVTQESGPAHRKEFTMTCRVERFIEIGSGTSKKLAK`
`RNAAAKMLLRVHTV`

The fact the very first residues of the constructs were not detected in HSQC spectra is very common in usual conditions because of the fast exchange with the solvent. All the ^{13}CO, ^{13}Cα and ^{13}Cβ resonances have been assigned apart for the first 3.

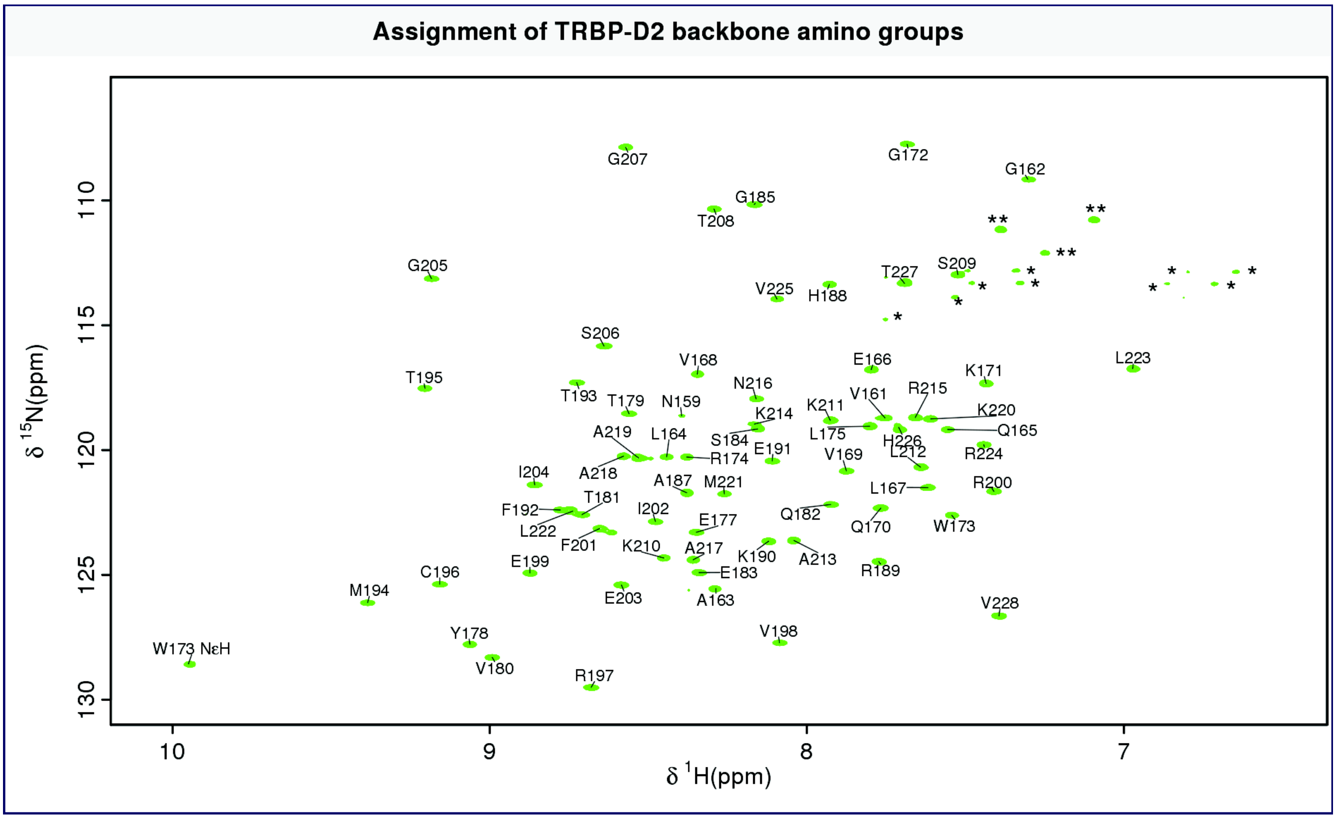


Figure 15.6: (^{1}H,^{15}N) BEST-TROSY spectrum of ^{15}N-labeled TRBP(157-228), recorded at 600 MHz and 25°C in [23.5 mM potassium phosphate pH 6.5, 100 mM KCl, 10 mM MgCl$_2$ and 5 mM β-mercaptoethanol, 5mM TCEP with 5% ^{2}H$_2$O added]. Backbone ^{15}N-^{1}H^{extN} resonance assignments are indicated as well as the signal coming from the tryptophan ϵ amino group. The unassigned signal peaks on the top-right corner labeled with stars (*) of the spectrum are generated by side chains amino groups; the three peaks of higher intensity in this region (**) have a spectrum width dependent position and are likely to be aliased arginine side chain amino groups.

15.4.2 Secondary structure and ^{1}H -^{15}N-NOE of TRBP-D2

Secondary structure The secondary structure was estimated using the ^{13}CO, ^{13}Cα and ^{13}Cβ chemical shifts using TALOS+ (247). The estimated secondary structure obtained with data collected from TRBP-D2 construct containing is in excellent agreement with the secondary structure found in the previously reported NMR structure (2CPN) of the second domain of TRBP (**fig. 15.7**).

^{1}H -^{15}N-NOE In order to characterize the flexibility of TRBP-D2, ^{1}H -^{15}N NOE were measured. This parameter provides information about the high frequency motions (10^{-8}- 10^{-12}

s^{-1}) of ^{1}H -^{15}N bond vectors (**fig. 15.7**). The NOEs were classically calculated (**section 12.2**). To avoid biases caused by chemical shift degeneracies, 2 residues closed to other peaks, S184 and H226, were not included in the analysis. The high ^{1}H -^{15}N NOE values in each dsRBD secondary structure elements confirmed that the second domain of TRBP alone is folded in the sampled time scale. One flexible region (resid. 184 to 191) is also detectable in the dsRBDs, corresponding the common β1-β2 flexible loop that has been observed in other dsRBDs.

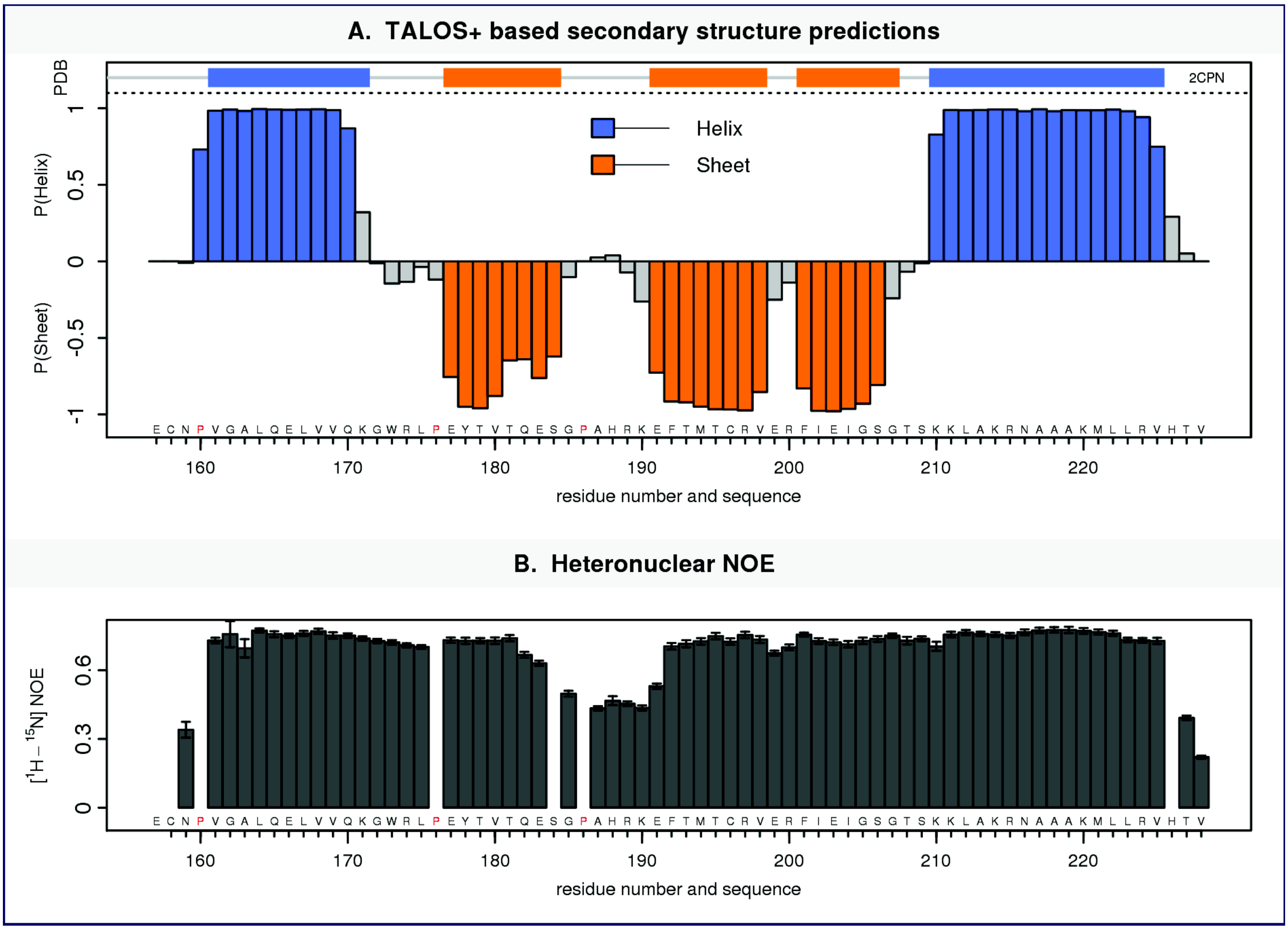


Figure 15.7: Secondary structure prediction based on TALOS+ (**A.**) and HN heteronuclear NOE (**B.**) for TRBP-D2 construct. The propensities in favor of mainly helical and sheet secondary fold are colored in blue and orange respectively. Error bar represents +/- one standard deviation estimated by simulating 100 ratios using the spectral noise and assuming a Gaussian distribution. The secondary structure obtained from the deposited structure of the second domain of TRBP (PDBid: 2CPN) is given for comparison with the same color code. For an easier visualization, the proline residues are represented in red.

15.5 NMR Assignment, secondary structure and dynamics of TRBP2-D1

HSQC spectra of TRBP-D1 The number of backbone amide groups detected in the spectra of TRBP2-D2 is globally equal to the non proline sequence length, apart from the three first undetected residues. On the contrary, at high signal to noise ratio, the HSQC or TROSY spectra of TRBP-D1 is reproducibly characterized by a total number of non side chain peaks slightly exceeding two times the number expected from the sequence. This is a reproducible feature, observed with different experiments (HSQC, TROSY) on different spectrometers and on every sample in the phosphate buffer with or without 5% D_20 added.

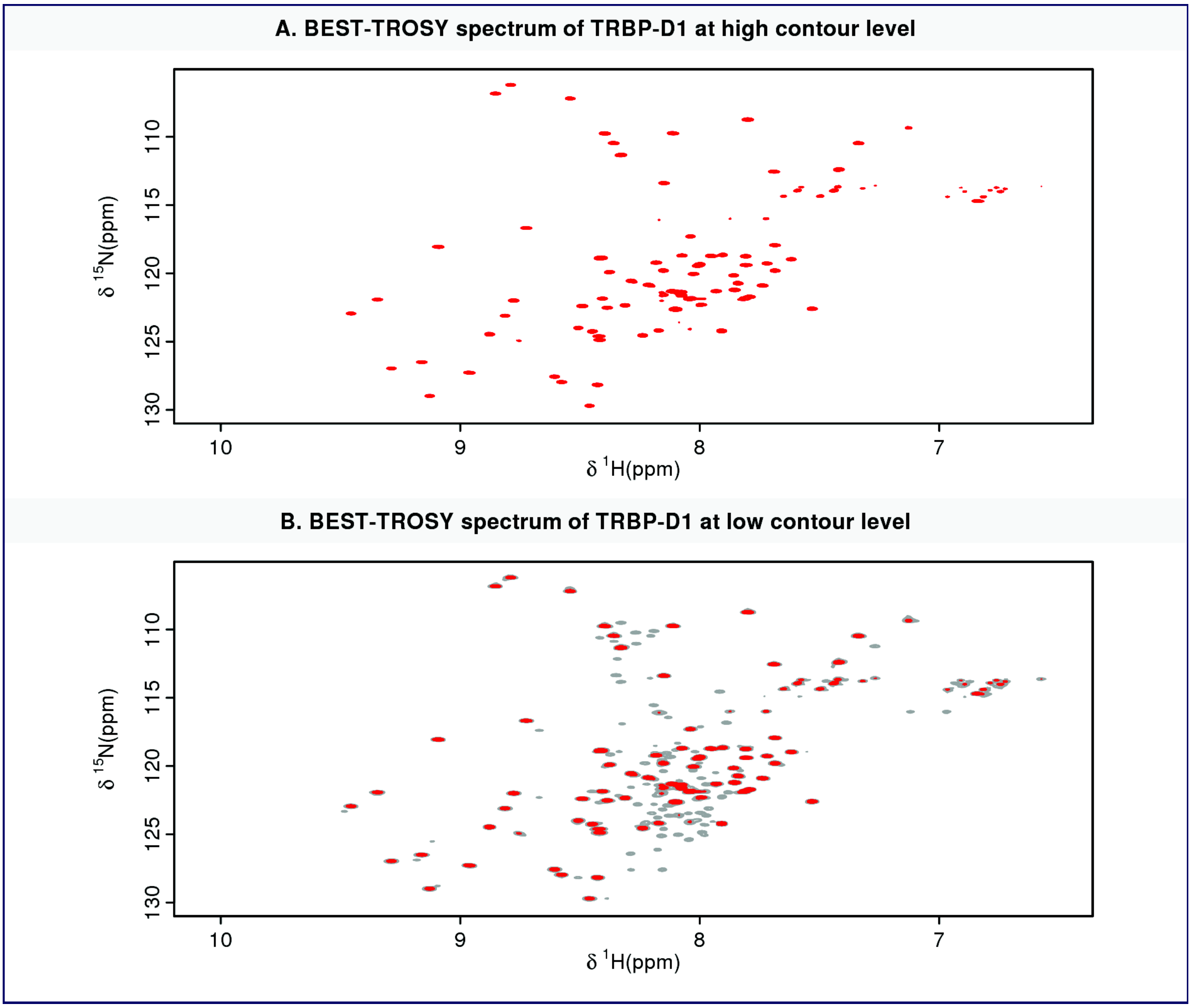


Figure 15.8: BEST-TROSY spectrum of TRBP-D1 (19-99) at 1.4 mM recorded at 600 MHz an 25°C in [23.5 mM potassium phosphate pH 6.5, 100 mM KCl, 10 mM $MgCl_2$ and 5 mM β-mercaptoethanol, 5mM TCEP, with 5% D_20 added]. The spectrum is plotted a high contour level (**A.**) or at a lower contour level with the superposition of the spectrum at high contour level (**B.**)

HSQC spectra of TRBP-D1 The dispersion of the peaks enable to distinguish artificially at least three sets of peaks (**fig. 15.8**) :

- The major set correspond to a well dispersed set of intense peaks similar to the one obtained for TRBP-D2. These are the peaks we see at high contour level in **fig. 15.8** A.

- The second set of peaks has about 20% the volume of the first one. A striking difference

between the first two sets of peaks is their chemical shift range. The set with the higher signal is well dispersed whereas this one is included in a narrow ^{1}H range (8 to 9 ppm) suggesting that the peaks come from an unfolded protein (**fig. 15.8** B).
- Additionally a few peaks are clearly doubled: they have a very similar NH chemical shift and very close CA, CB and CO chemical shifts **fig. 15.8** B. This adds another complexity to the system.

Deeper investigation of these forms is presented after the establishment of a similar behavior in TRBP-D1D2.

15.5.1 Assignment strategy of the set of intense peaks in TRBP-D1 spectrum

Experiments and results As for TRBP2-D2, a set of BEST experiments has been used to assign the [^{1}H,^{15}N] TRBP2-D1 resonances: HN(CO)CACB, HNCACB, HNCO and inter-residual HADAMAC experiments. Correlations between the spins related to the peaks of high intensity have shown that these peaks belong to TRBP-D1 (**fig. 15.9**). Thus there is a least two species in the sample: TRBP-D1 and a probable unfolded species. Because of the abundance of the latter form, about 1/5 of the main one, and since only one band is seen in SDS-PAGE gel, the second set of peaks must be another species of TRBP-D1. By now, I will call the main species TRBP-D1^F and the main minor one TRBP-D1^U

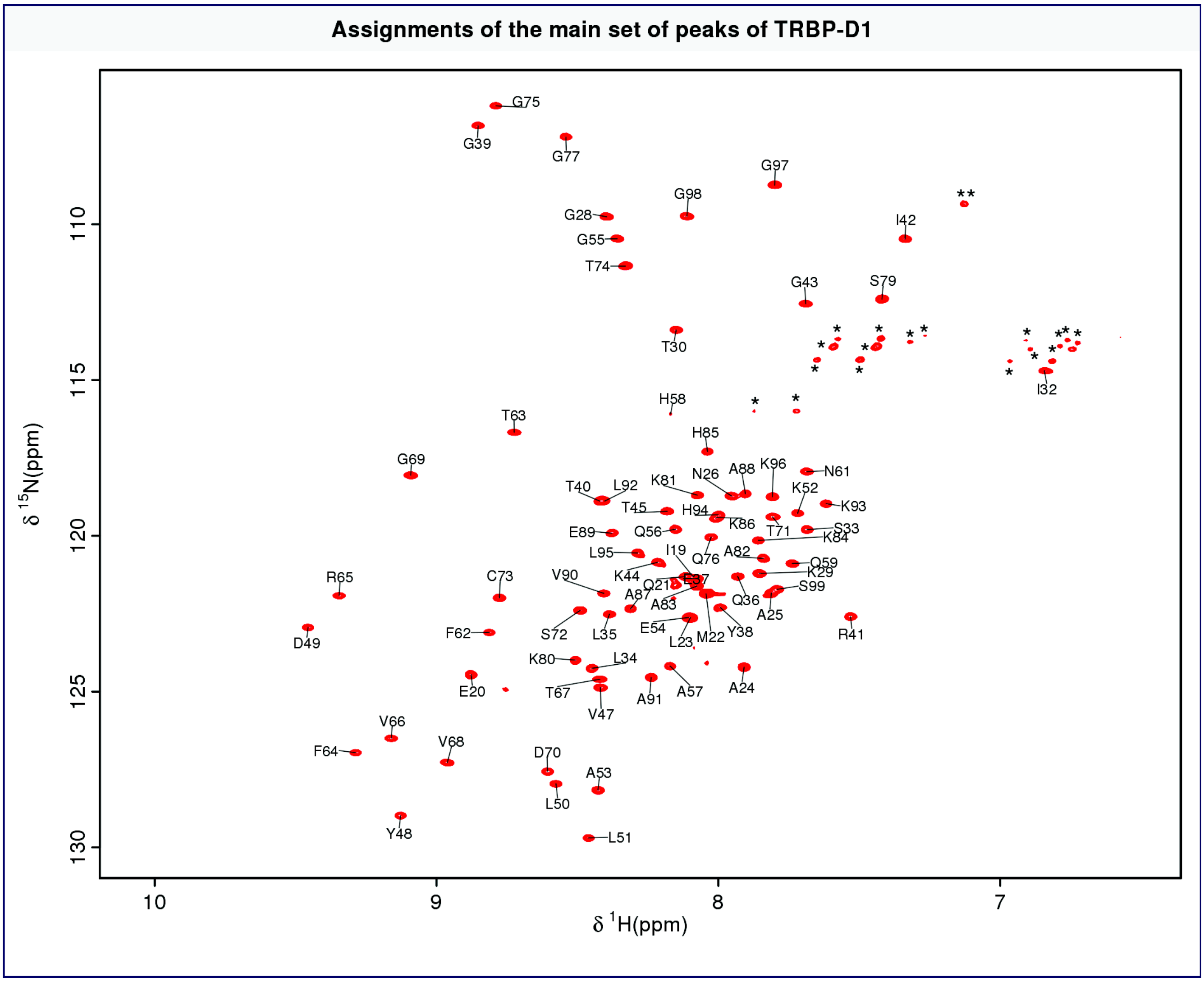

Assignments of the main set of peaks of TRBP-D1

Figure 15.9: (^{1}H,^{15}N) BEST-TROSY spectrum of ^{15}N-labeled TRBP-D1 (19-99), recorded at 600 MHz and 25°C in [23.5 mM potassium phosphate pH 6.5, 100 mM KCl, 10 mM MgCl$_2$ and 5 mM β-mercaptoethanol, 5mM TCEP with 5% D$_2$0 added]. Backbone ^{15}N-^{1}H^{extN} resonance assignments are indicated. The unassigned signal peaks on the top-right corner of the spectrum are generated by side chains amino groups and are labeled with stars (*). A peak of higher intensity in this region (**) has a spectrum width dependent position and is likely to be aliased arginine side chain amino groups.

89% of the backbone [^{1}H,^{15}N] resonances have been assigned; the missing residues are the three first ones including the residual glycine from the purification tag. The residues without [^{1}H,^{15}N] resonances assigned are represented in red in the following sequence of the construct and the prolines with no amino groups in grey :

`GIEQMLAANPGKTPISLLQEYGTRIGKTPVYDLLKAEGQAHQPNFTFRVTVGDTSCTGQG`

`PSKKAAKHKAAEVALKHLKGGS`

All the ^{13}CO, ^{13}Cα and ^{13}Cβ resonances have been assigned apart for the first 4 residues.

15.5.2 Secondary structure and ^{1}H -^{15}N-NOE of TRBP-D1^{F}

Secondary structure The secondary structure of TRBP-D1^{F} was estimated with the same methods as TRBP-D2. It is very similar to the one of TRBP-D2 (**fig. 15.10**). The estimated secondary structure is also in excellent agreement with the secondary structure found in the previously reported crystal structure (PDBid: 3LLH) of the first domain of TRBP.

^{1}H -^{15}N-NOE In order to characterize the flexibility of TRBP-D1^{F}, ^{1}H -^{15}N NOE were measured (**section 12.2**), and the results are given in **fig. 15.10**. The profile is highly similar although the lower NOE values in the β1-β2 loop suggests more flexibility of this region in this domain, which could be linked to two more residues in this loop compared to TRBP-D2

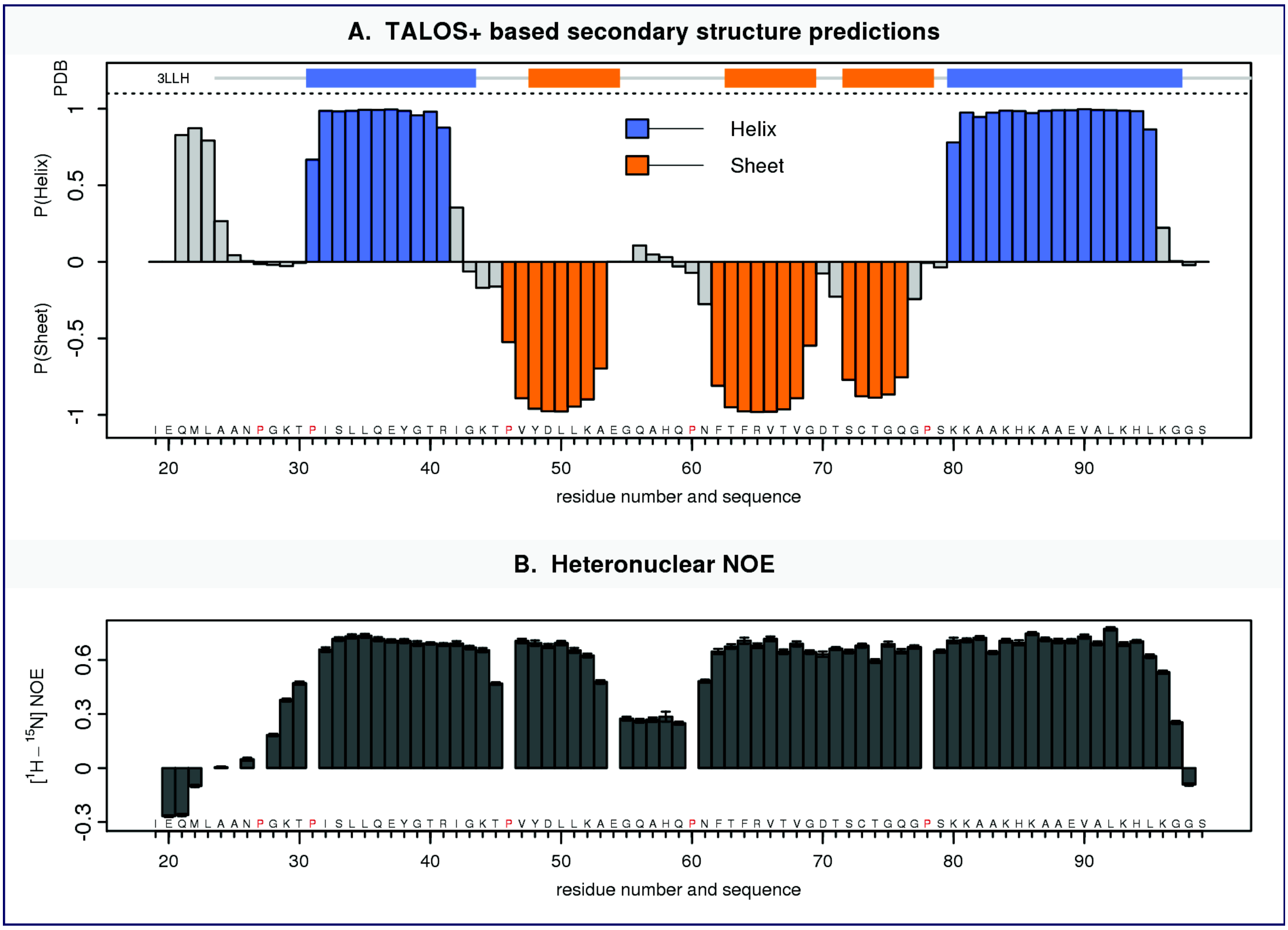
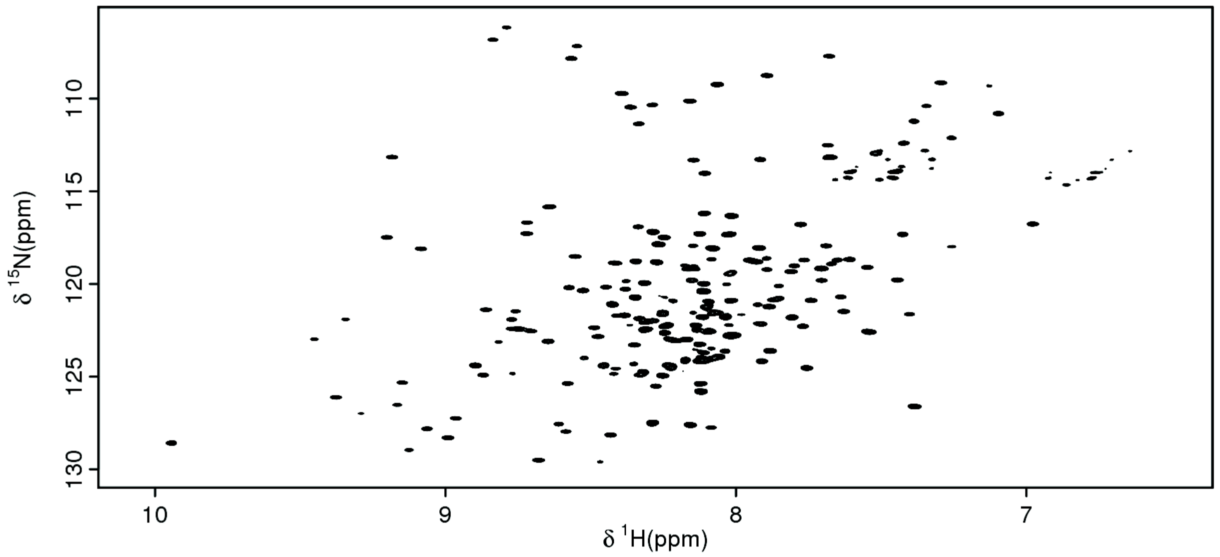


Figure 15.10: Secondary structure prediction based on TALOS+ (**A.**) and HN heteronuclear NOE (**B.**) for the intense set of peaks of TRBP-D1 construct. The propensities in favor of mainly helical and sheet secondary fold are colored in blue and orange respectively. The secondary structure obtained from the deposited structure of the second domain of TRBP (PDBid: 3LLH) is given for comparison with the same color code. For an easier visualization, the proline residues are represented in red. Error bar represents +/- one standard deviation estimated by simulating 100 ratios using the spectral noise and assuming a Gaussian distribution.

15.6 NMR Assignment, secondary structure and dynamics of TRBP-D1D2

HSQC spectra of TRBP-D1D2 (19-288) As for TRBP-D1, various sets of peaks are reproducibly visible on the spectra of TRBP-D1D2 (19-228). The number of non side chain amino peaks exceeds 300 and includes about 160 ones in the central region H 7.75 ppm - 8.5 ppm (**fig. 15.11**). The same naming convention (F/U) is use to refer to the main form (TRBP-D1^{F}D2) and the form with an unfolded state (TRBP-D1^{U}D2).

A. BEST-TROSY spectrum of TRBP-D1D2 at high contour level



B. BEST-TROSY spectrum of TRBP-D1D2 at low contour level

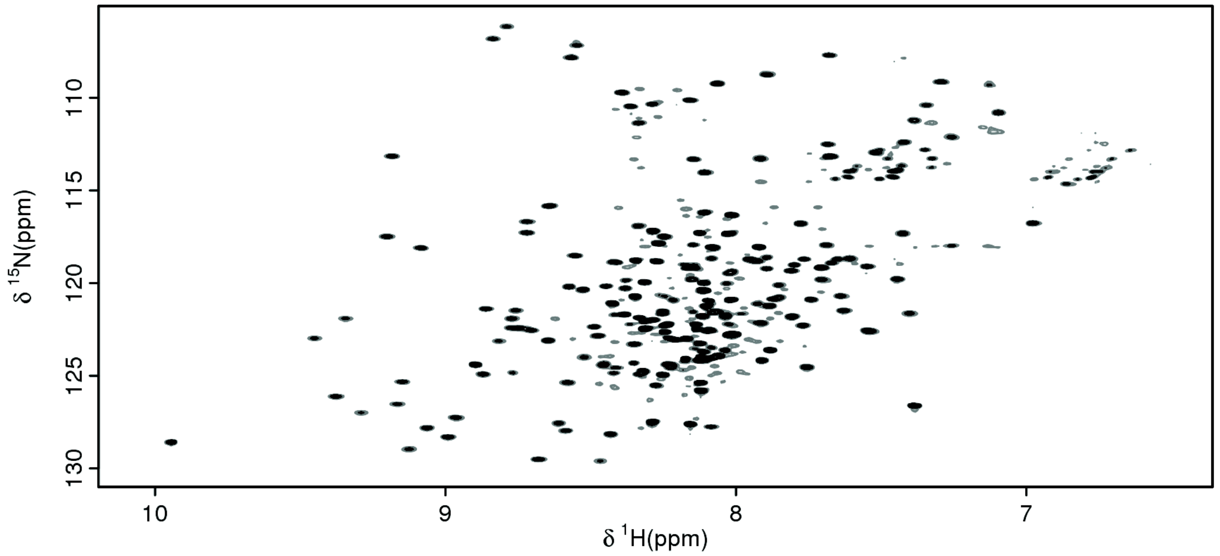


Figure 15.11: BEST-TROSY spectrum of TRBP-D1D2 (19-228) at 1.4 mM recorded at 600 MHz and 25°C in [23.5 mM potassium phosphate pH 6.5, 100 mM KCl, 10 mM $MgCl_2$ and 5 mM β-mercaptoethanol, 5mM TCEP, with 5% D_20 added]. The spectrum is plotted a high contour level (**A.**) or at a lower contour level with the superposition of the spectrum at high contour level (**B.**)

Table 15.3: Main parameters of the NMR experiments recorded on TRBP(19-228). t^{max}: maximum acquisition time

Sample	Experiment	Recycle delay(s)	number of scans	t^{max}(H) (ms)	t^{max}(N) (ms)	t^{max}(C) (ms)
protonated	BEST-TROSY-HSQC	0.5	120	70	80	na
deuterated	BEST-TROSY-HSQC	1.5	8	70	75	na
deuterated	BEST-TROSY-HN(CO)CACB	1.5	8	64	26	8.3 (CACB)
deuterated	BEST-TROSY-HNCACB	1.5	8	64	24	8.3 (CACB)
deuterated	BEST-TROSY-HNCO	1.5	4	64	26	10 (CO)
protonated	^{1}H -^{15}N-het-NOE	5	24	70	81	na

15.6.1 Backbone assignment of TRBP-D1^{F}D2

Experiments Because of the larger size of TRBP2-D1D2(19-228) (22.6 kDa) compared to the single domains, the fact the spectrum of the first domain was already crowded and in order to prepare the analysis of the interaction with RNA, the assignment of TRBP-D1D2 has been performed on a sample of produced in a *U*-[^{2}H,^{13}C,^{15}N]-labeled media. The deuterated protein was obtained using the double colony selection protocol (**section 13.4.1**) and purified in H_2O to allow exchange of the amide deuterons with protons. Three dimensional NMR experiments for backbone assignment were recorded using 90 μL of a 600 μM sample of TRBP-D1D2 (19-228) in a 2.5 mm diameter Shigemi tube (Shigemi Inc.) placed in a 5 mm diameter standard tube filled with D_2O. 3D BEST-TROSY versions of HNCACB, HNCOCACB and HNCO experiments (77) were performed at 25°C using a Varian (Agilent) DirectDrive 600 MHz spectrometer equipped with a triple resonance cryogenic probe. As this experiments concerns the whole RNA binding region of TRBP, acquisition parameters are detailed in **tab. 15.3**

Results and analysis The assignment was performed without using the assignment of the single domains. However, the analysis of the belonging of the peaks to each domain was regularly checked for consistency with an overlay of the spectra of TRBP-D1, TRBP-D2 and TRBP-D1D2. The assignment is presented on a protonated sample [^{1}H,^{15}N] to facilitate the comparison with the assignment of the single domains **fig. 15.12**. It was possible to obtain at least one backbone nuclear assignments (H,N,C) for 209 out of 210 residues of TRBP-D1D2 (19-228), including 99.5% of non proline ^{15}N-H^N pairs and $^{13}C^\alpha$, $^{13}C^\beta$ chemical shifts, 91% of $^{13}C'$ chemical shifts and 98.5 % of backbone amino groups. Unassigned backbone amino groups are represented in red in the following sequence and the proline in gray :

```
GIEQMLAANPGKTPISLLQEYGTRIGKTPVYDLLKAEGQAHQPNFTFRVTVGDTSCTGQGPSKKAAKHKAA
EVALKHLKGGSMLEPALEDSSSFSPLDSSLPEDIPVFTAAAAATPVPSVVLTRSPAMELQPPVSPQQSECN
PVGALQELVVQKGWRLPEYTVTQESGPAHRKEFTMTCRVERFIEIGSGTSKKLAKRNAAAKMLLRVHTV
```

Resonance assignments of TRBP-D1D2 (19-228) have been deposited in the BioMagResBank (BMRB; http://www.bmrb.wisc.edu) under the BMRB accession number 18324. This assignement of TRBP-D1D2 has been the subject of an assignment note (16)

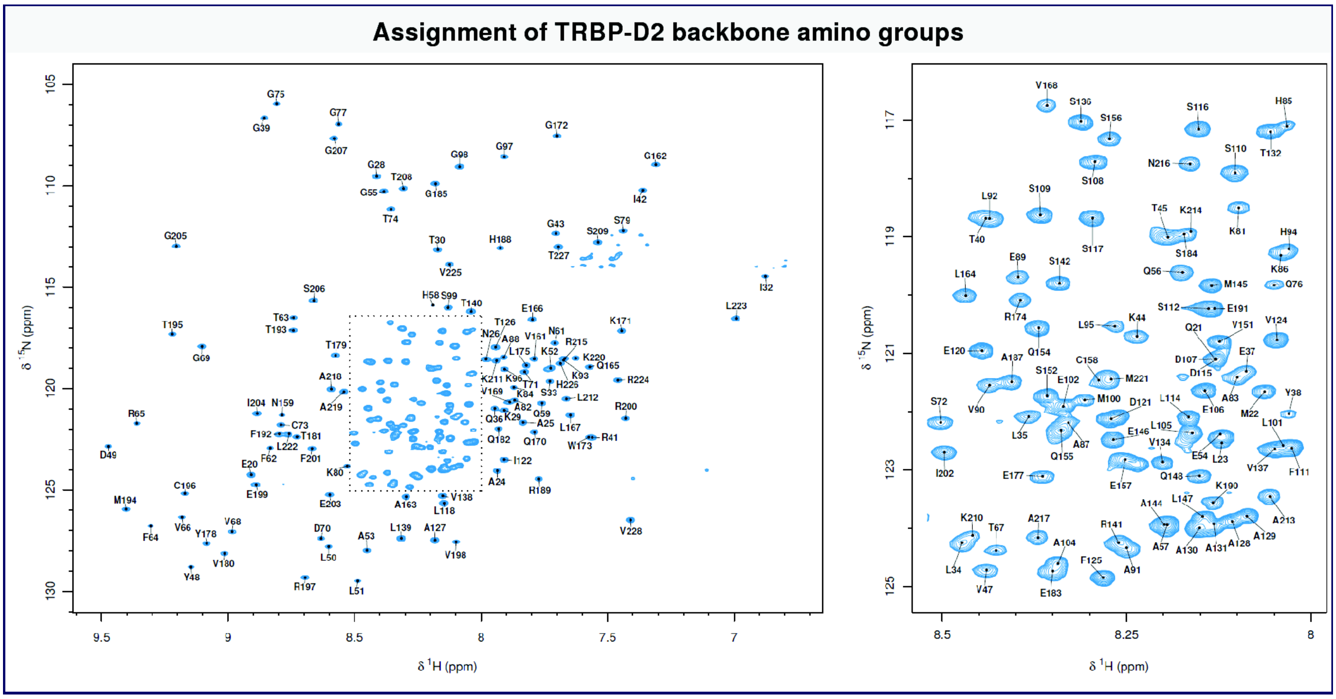


Figure 15.12: (^{1}H,^{15}N) TROSY-HSQC spectrum of ^{15}N-labeled TRBP-D1D2 (19-228) with backbone ^{15}N-^{1}H^N resonance assignments recorded at 800 MHz and 25°C in [23.5 mM potassium phosphate pH 6.5, 100 mM KCl, 10 mM MgCl$_2$ and 5 mM β-mercaptoethanol, 2 mM TCEP]. The panel on the right shows an expansion of the region of the full spectrum contained in the dashed box.

15.6.2 Secondary structure and ^{1}H -^{15}N-NOE of TRBP-D1^{F}D2

Secondary structure The secondary structure was estimated with the same method than TRBP-D2, using TALOS+ (247). The secondary structure prediction of TRBP-D1D2 is in excellent agreement with the previously reported 3D structures of the two individual domains (290) (PDBid: 2CPN, 3LLH) (**fig. 15.13**).

^{1}H -^{15}N-NOE In order to characterize the flexibility of TRBP-D1D2, ^{1}H -^{15}N NOE were measured. Heteronuclear ^{1}H-^{15}N NOE experiments (74) were measured on a Varian (Agilent) DirectDrive 600 MHz at 25 °C with a room temperature probe using 300 μL of a 1.4 mM sample of [^{15}N]-labeled TRBP(19-228) in a 5 mm Shigemi tube (Shigemi Inc.) with 5% D$_2$O. The main parameters used for the experiment are given in **tab. 15.3**. The NOEs were calculated as described in **section 12.2**. To avoid bias caused by chemical shift degeneracies, 29 peaks with close ^{1}H -^{15}N chemical shifts were not included in the analysis (**fig. 15.13**). The high ^{1}H-^{15}N NOE values in the secondary structure elements of both domains confirm that the two dsRBDs are folded like in the single domain studies. The interdomain region connecting the two dsRBDs (residues 99 to 159), without significant secondary structure propensities, is characterized by negative heteronuclear NOE values indicating a high degree of flexibility; it is thus a flexible linker.

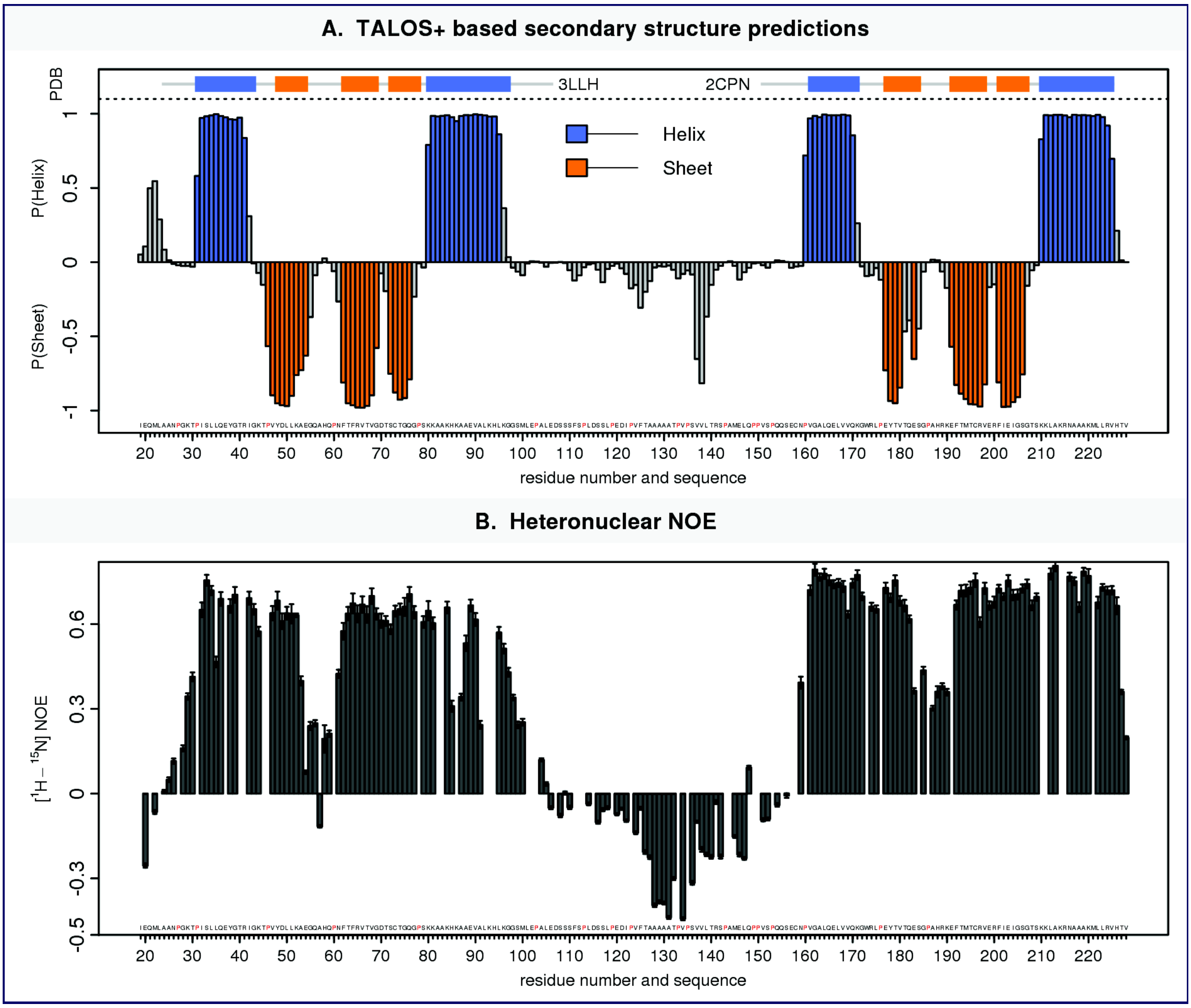


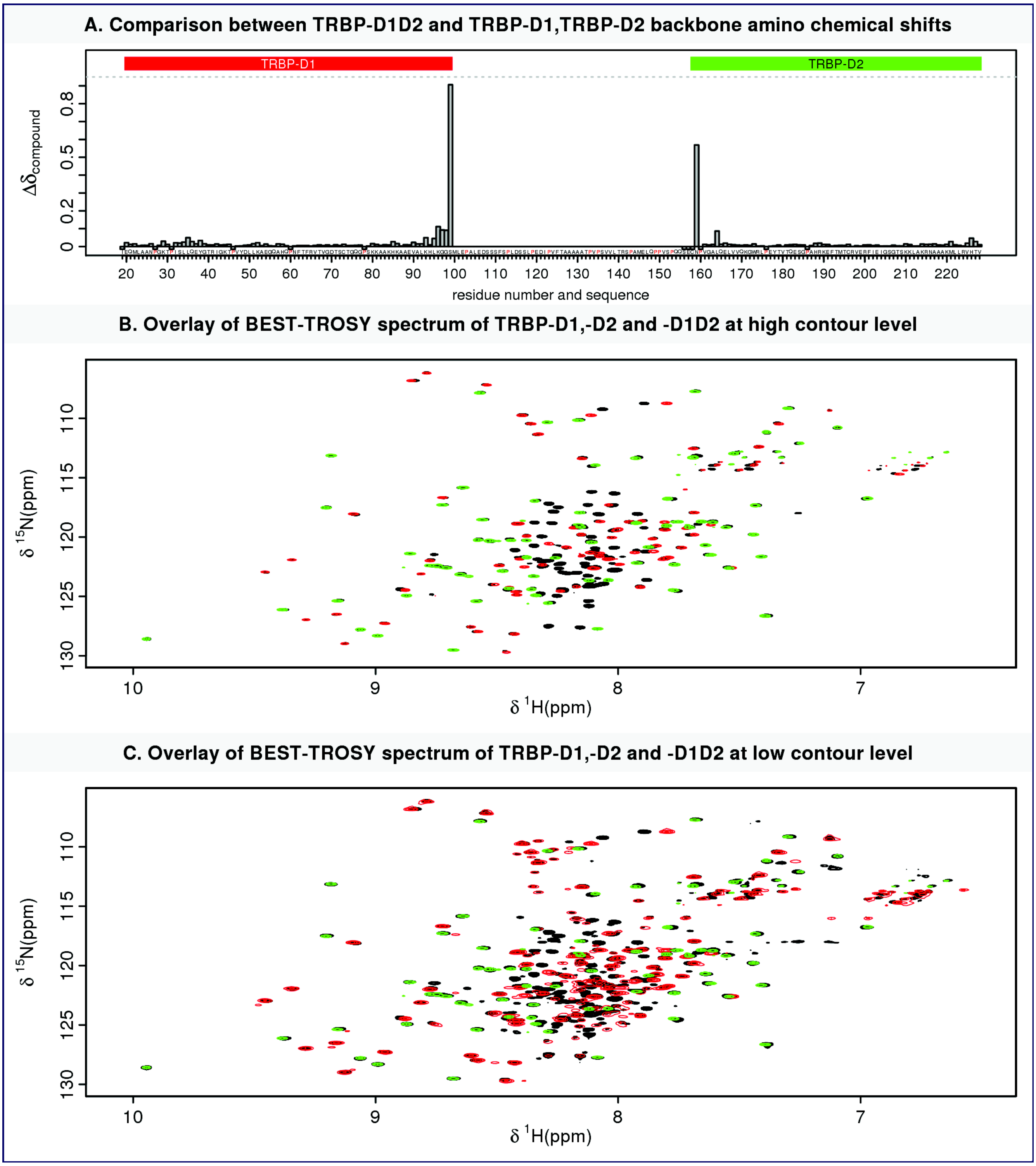
Figure 15.13: Secondary structure prediction based on TALOS+ (A.) and HN heteronuclear NOE (B.) for TRBP-D1D2 construct. The propensities in favor of mainly helical and sheet secondary fold are colored in blue and orange respectively. The secondary structure obtained from the deposited structure of the isolated dsRBDs of TRBP (PDBid: 2CPN, 3LLH) is given for comparison with the same color code. For an easier visualization, the proline residues are represented in red. Error bar represents +/- one standard deviation estimated by simulating 100 ratios using the spectral noise and assuming a Gaussian distribution.

The only positive heteronuclear NOE value in the linker region concerns a residue preceding two adjacent prolines, a motif that is known to add local order to polypeptide chains. Finally, each domain contains a short stretch of decreased heteronuclear NOE values that centers on the conserved histidine of the loop between β-strands 1 and 2 (residues 53-61, dsRBD-1, and residues 183-191, dsRBD-2). This loop comprises the second of three regions in dsRBDs that form direct or water-mediated contacts with dsRNAs (264). The enhanced flexibility of this loop compared to the rest of the domain may indicate a role in target recognition.

15.7 Comparison between apo TRBP-D1D2 and TRBP-D1 & TRBP-D2

15.7.1 Chemical shifts comparison between TRBP-D1D2 and TRBP-D1 & TRBP-D2

The absence of intermolecular interaction in the samples of TRBP-D1D2, TRBP-D1 and TRBP-D2 has been established previously by SEC-MALLS, complemented with svAUC for TRBP-D1D2. In order to investigate potential intramolecular interaction between the first two dsRBDs of TRBP and/or between the dsRBD and the interdomain region, a comparison of the chemical shifts of TRBP-D1^{F}D2 with TRBP-D1^{F} & TRBP-D2 was performed in **fig. 15.14**. It represents the variation of chemical shift in TRBP-D1 and TRBP-D2 with TRBP-D1D2 as a reference. We see that all shifts are very small (< 0.05 ppm) apart a few residues located at the C-terminus of TRBP-D1 and at the N-terminus of TRBP-D2. The latter changes are expected because of the large change of chemical environment related to the presence of the interdomain region, including 2 new covalent bounds at each extremity of the linker region. The result shows that apart from these trivial 3 residues, the chemical environments of the two first domains of TRBP are the same in the tandem construct or when they are isolated. This imply that there are no significant intramolecular interaction between the domains influencing chemical shifts, neither interactions between the linker and the domains, apart from the localized truncation effects on the linker boundaries. TRBP-D1^{U} has also an extremely similar spectrum in TRBP-D1 and TRBP-D1D2 samples, which is similar to what is observed for the folded form of D1. (**fig. 15.14**, C.). Thus the two first domain of TRBP have an independent conformation in their apo form. An overlay of the spectra of TRBP-D1,TRBP-D2 and TRBP-D1D2 show differently this independency (**fig. 15.14**). The assignment of each peak, required for a proper comparison is not plotted for an easier visualization since the identicity of the closest peak information is already contained in **fig. 15.14**.


A. Comparison between TRBP-D1D2 and TRBP-D1,TRBP-D2 backbone amino chemical shifts

B. Overlay of BEST-TROSY spectrum of TRBP-D1,-D2 and -D1D2 at high contour level

C. Overlay of BEST-TROSY spectrum of TRBP-D1,-D2 and -D1D2 at low contour level

Figure 15.14: Comparison of the chemical shifts of TRBP-D1,TRBP-D2 and TRBP-D1D2. **A.** Estimation of the deviation of backbone H-N chemical shifts for the residues assigned in both TRBP-D1^{F}, TRBP-D2 and TRBP-D1^{F}D2 in their apo forms. Chemical shifts deviation of ^{1}H and ^{15}N nucleus are composed in a combined "compound" chemical shift. Unassigned residues and prolines are indicated by a small negative bar. **B., C.** Overlay of (^{1}H,^{15}N) BEST-TROSY spectrum of ^{15}N-labeled TRBP-D1(19-99), TRBP-D2(157-228) and TRBP-D1D2(19-228) at 1.4, 1.4 and 1 mM respectively, recorded at 600 MHz and 25°C in [23.5 mM potassium phosphate pH 6.5, 100 mM KCl, 10 mM $MgCl_2$ and 5 mM β-mercaptoethanol, 5mM TCEP with 5% D_20 added]. Less coutours are used for the single domain constructs (TRBP-D1 and TRBP-D2) than for TRBP-D1D2 to be able to see the peaks overlay. The comparison is plotted at high contour level (**A.**) or low contour level (**B.**)

15.7.2 Comparison of the secondary structure and Het-NOE between TRBP-D1^{F}D2 and TRBP-D1^F & TRBP-D2

Secondary structure and ^{1}H-^{15}N NOE comparison To further get insight on the behavior of the first two dsRBDs of TRBP within the miRNA binding region, the results of secondary structure prediction and ^{1}H-^{15}N obtained previously were compared for TRBP-D1^{F}D2 with TRBP-D1^F and TRBP-D2 **fig. 15.15**. The secondary structure and ^{1}H-^{15}N NOE match very well in the isolated domain and in the tandem construct. Complementary to the previous ^{1}H-^{15}N chemical shifts comparison, these observations confirm the independency of the first two dsRBDs of TRBP in solution. On average, ^{1}H-^{15}N NOE values are slightly lower in the first dsRBD compared to the second one, either in the tandem construct or in isolated domains. This could be related to a difference of stability between the fold of the two dsRBDs. The next paragraphs explore the special behavior of the first dsRBD of TRBP. Additionally, the β1-β2 loop of the first dsRBD has lower ^{1}H-^{15}N NOE values compared to the second dsRBD in the tandem construct and in isolated domains. This may just relates to the fact the loop is slightly longer in the first dsRBD (8 residues) than in the second one (6 residues).

15.8 Analysis of TRBP-D1^U

Efforts have been directed to characterize TRBP-D1^U. Since this minor species has very similar HN chemical shifts in TRBP-D1 and TRBP-D1D2 spectra and because of the similarity of the behavior of the main species of the first domain of TRBP in isolated or in the tandem construct (HN chemical shift, secondary structure and HN-NOE) it is reasonable to consider that the same phenomenon is responsible for the existence of the minor species in both samples. This is why the construct TRBP-D1, with less signal crowding than TRBP-D1D2 has been used to analyze TRBP-D1^U. The [^{1}H,^{15}N] chemical shift dispersion of species U is characteristic of unfolded proteins. Taking into account TRBP-D1 has be shown to be monomeric in the previous chapter, two hypotheses can be proposed to explain the presence of this species :

- Independent signals. This would mean that there is a folded species and an independent unfolded species: either part of the folded protein unfolds in the production process, either two forms, folded and unfolded, are purified from *E. coli* cells.
- Dependent signals. Both signals come from the same polypeptide, meaning that it undergoes an exchange between the two detected sets of peaks. This would imply that the protein alternates between a folded and unfolded conformation.

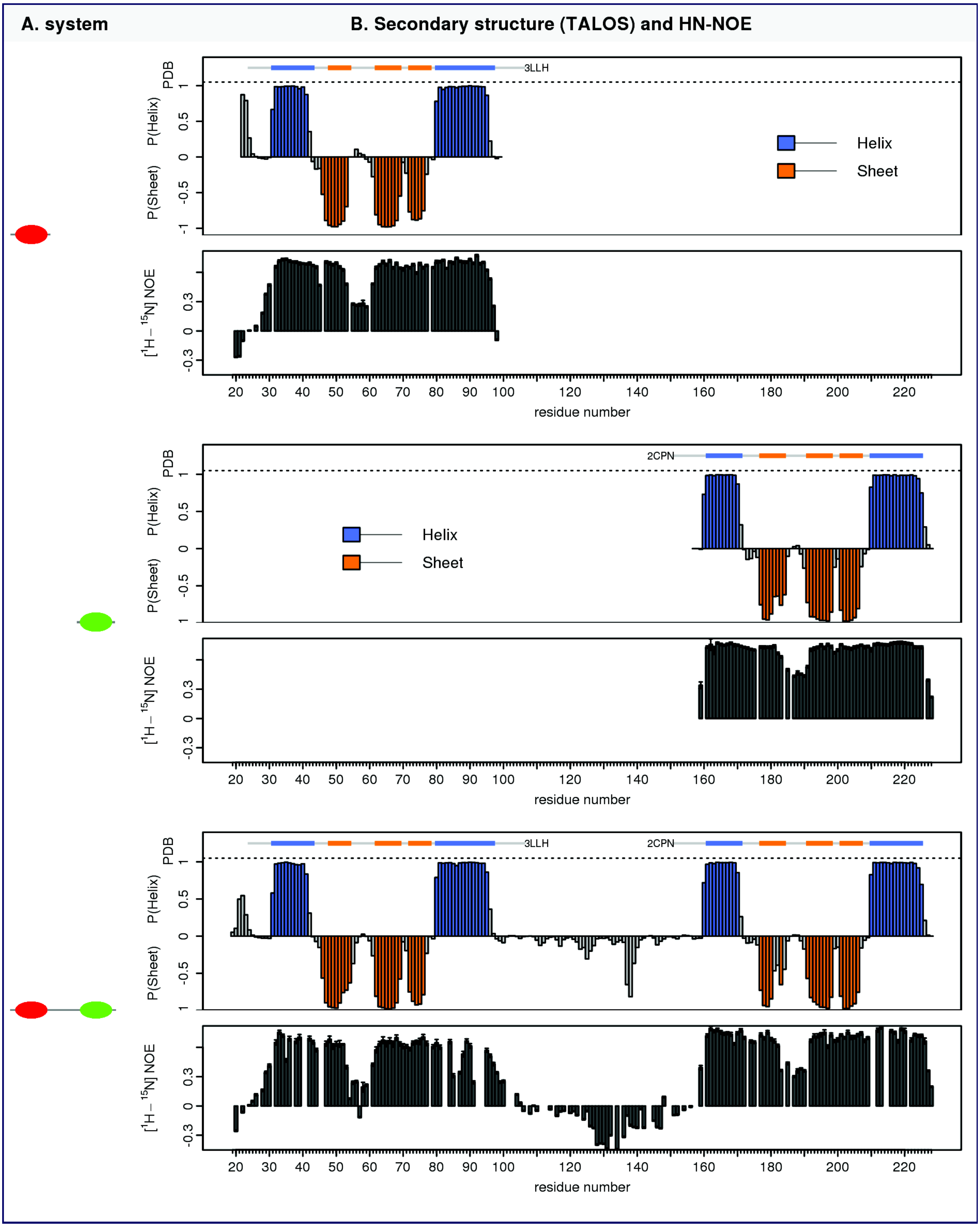


Figure 15.15: Comparison of the secondary structures prediction based on TALOS+ and HN heteronuclear NOE for TRBP-D1, TRBP-D2 and TRBP-D1D2 construct. The propensities in favor of mainly helical and sheet secondary fold are colored in blue and orange respectively. The secondary structures obtained from the deposited structure of the first and second domain of TRBP (3LLH and 2CPN respectively) are given for comparison with the same color code. Error bar represents +/- one standard deviation estimated by simulating 100 ratios using the spectral noise and assuming a Gaussian distribution.

15.8.1 SEC analysis of TRBP-D1

Previous knowledge of TRBP-D1 profile on SEC was first used to clarify the situation. It is known that folded and unfolded form of one protein have different elution volume in an appropriate size exclusion chromatography for the protein under study. They elute as molecules of molecular weight 2 to 6 times higher (271; 269) than they folded counterpart. So if the first hypothesis is true, with two distinct forms of the proteins before the SEC step, then two peaks are expected in the chromatogram, or at least a peak much broader than the one of TRBP-D2. However only one was reproducibly detected, with the same shape than the one of TRBP-D2. It was still possible that partial unfolding of TRBP-D1 occur during the concentration step often used after the SEC step to prepare the samples for NMR, or more generally between the SEC step and the NMR analysis. However, the spectra of TRBP-D1 recorded just after the SEC and without concentration are also characterized by the two sets of peaks. Additionally the chromatogram of a sample after NMR analysis is also characterized by one peak. So the unfolded form was not induced by the post-SEC steps. These results were not consistent with the first hypothesis and led me to the formulation of the second alternative hypothesis: the folded and unfolded species of TRBP are two forms of the same protein molecule. Since one set of peaks per species are detected, NMR was an excellent technique to test the latter hypothesis. If two forms are in exchange in the millisecond to second time scale, the Exchange SpectroscopY technique (EXSY) can detect the correlation between the peaks of the two forms, as explained **section 12.3.3.1**.

15.8.2 BEST-EXSY of TRBP-D1: evidence of exchange

In order to detect potential exchange between TRBP-D1^F and TRBP-D1^U, some BEST-EXSY experiments have been performed: one with a mixing time of 0 s and another with a mixing time of 0.16 s. The results of these experiments are given in **fig. 15.16**. The spectra show that each peak of the main species of TRBP-D1 (in red on the figure) is in exchange with a peak of the species 2. Thus these two species are actually two forms in exchange. No exchange is observed with the peaks of the third species.

15.8.3 Quantitative NMR analysis of TRBP-D1^U

The narrow chemical shift range of the form U already suggested that the form was an unfolded one. But to get more information on this form including the residues affected by the exchange, a direct proof of the unfolded state and the potential residual secondary structure, additional NMR experiments have been used.

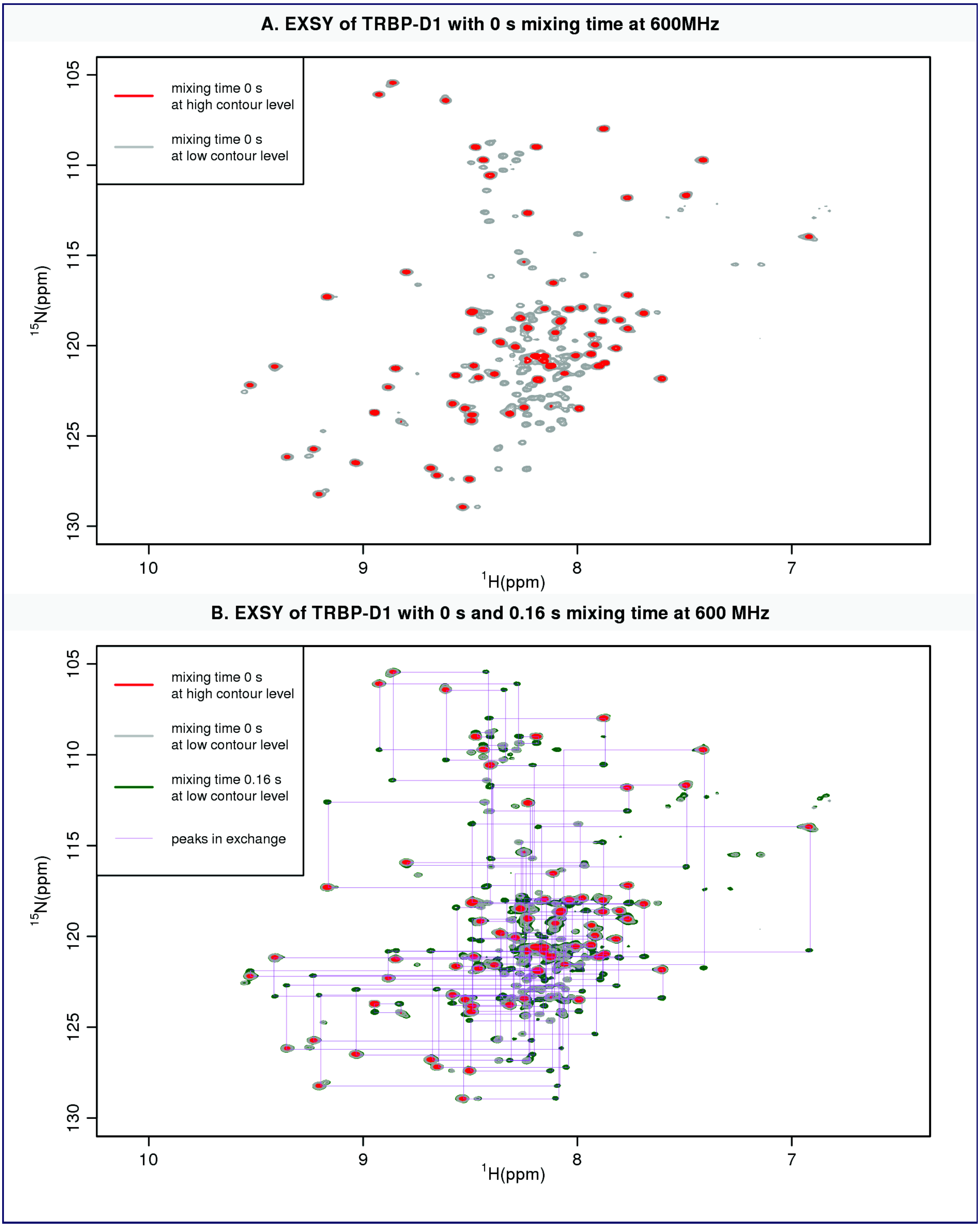


Figure 15.16: BEST-EXSY experiments of TRBP-D1 with a 0 s mixing time and 0.16 s mixing time recorded at 600 MHz on a sample at 600 μM and 25°C in [23.5 mM potassium phosphate pH 6.5, 100 mM KCl, 10 mM $MgCl_2$ and 5 mM β-mercaptoethanol, 5mM TCEP]. **A.** Spectra with 0 s mixing time. To easily see the main species of TRBP-D1 compared to the other species, the spectrum is plotted at high and low contour level. **B.** Spectra with 0.16 s mixing time, overlaid with the spectra A. The identified peaks in exchange are linked by rectangle purple lines.

15.8.3.1 Assignment of TRBP-D1^F

In order to analyze the second form by NMR, the first step was to assign its ^{1}H, ^{15}N, Cα, Cβ and CO resonances. These experiments were recorded on a 300 μl sample of TRBP-D1 at 1.4 mM in the phosphate buffer with 5% ^2H_2O added and placed in a 5 mm diameter Shigemi tube. Three types of experiments were used to achieve this using a 600 MHz spectrometer: the set of 3D HNC experiments used to assign the main form of TRBP, a (2D) EXSY experiment with a mixing time of 0.16 s to transfer the assignment from the peaks of the main form to the one of the second form, and a 3D EXSY experiment with a mixing time of 0.16 s that was very useful to help the assignment in the crowded central region of the spectrum. The assignment result is given in **fig. 15.17** together with the assignment of the main form.

15.8.3.2 Secondary structure and ^{1}H - ^{15}N-NOE of TRBP-D1^F

Using the same protocol as for the one used for TRBP-D1^F, the secondary structure propensities and ^{1}H - ^{15}N-NOE were estimated (**fig. 15.18**). The differences between the chemical shift of the Cβ and Cα chemical shifts is also provided since it is well known to depends on the secondary structure. No clear secondary structure propensities are detected in this analysis suggesting that the U form is actually really unstructured. This property seems incompatible with the definition of a molten globule state. The ^{1}H - ^{15}N-NOE are low, globally positive apart from the boundaries of the construct, with a mean value of about 0.4. The ^{1}H - ^{15}N-NOE of TRBP-D1^F are given for comparison in **fig. 15.18**. These results indicate that TRBP-D1^U is clearly less structured and more flexible than TRBP-D1^F.

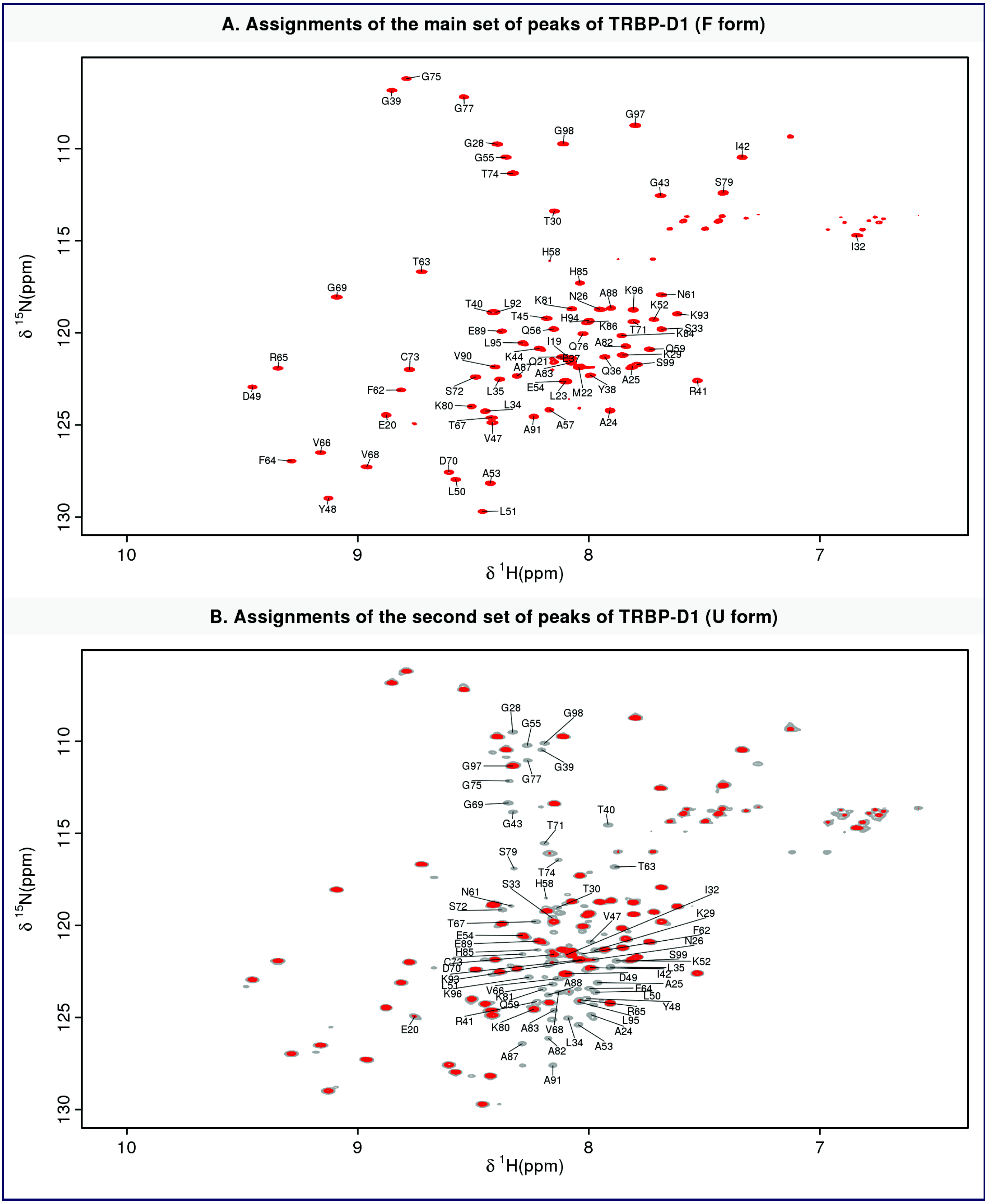


Figure 15.17: (^{1}H,^{15}N) BEST-TROSY spectrum of ^{15}N-labeled TRBP-D1(19-99), recorded at 600 MHz and 25°C in [23.5 mM potassium phosphate pH 6.5, 100 mM KCl, 10 mM $MgCl_2$ and 5 mM β-mercaptoethanol, 5mM TCEP with 5% D_20 added]. Backbone ^{15}N-^{1}H^{extN} resonance assignments are indicated for the major of minor form. The unassigned signal peaks on the top-right corner of the spectrum are generated by NH side chains.

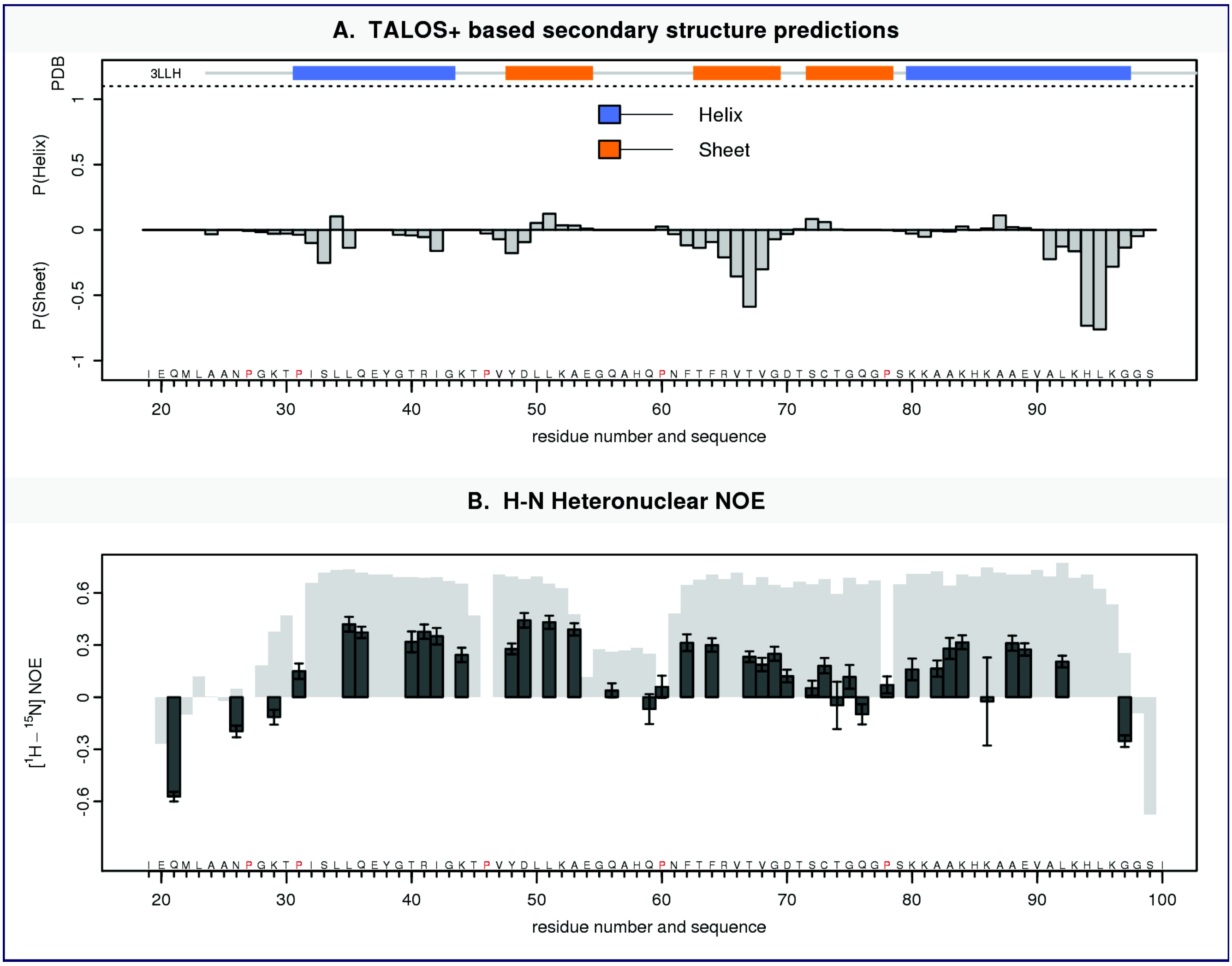


Figure 15.18: Secondary structure prediction based on TALOS+ (**A.**) and ^{1}H - ^{15}N-NOE (**B.**) for TRBP-D1^{U}. **A.** The secondary structure obtained from the deposited structure of the first domain of TRBP (PDBid: 3LLH) is given for comparison with the same color code. For an easier visualization, the proline residues are represented in red in the sequence. **B.** ^{1}H - ^{15}N-NOE of TRBP-D1^{U} in dark gray, overlaid n the ones of TRBP-D1^{F} plotted in light gray. Error bar represents +/- one standard deviation estimated by simulating 100 ratios using the spectral noise and assuming a Gaussian distribution.

15.8.3.3 Relative populations of TRBP-D1^{F} and TRBP-D1^{U}

The assignment of both TRBP-D1^{F} and TRBP-D1^{U} enables to compare the populations of the two states using relative volumes of the well resolved peaks for both states in fully relaxed NMR experiment (**fig. 15.19**). The mean of the volumes of well resolved peaks of the U form over the F form is about 0.21 with a standard deviation of 0.06, which is rather high and corresponds to a strong contribution of the noise in the spectrum and some residue related heterogeneity. Neglecting the few peaks found in TRBP-D1 and not in TRBP-D1D2, TRBP-D1^{F} alternates between a least two states folded (F) and unfolded (U) of average populations 83% and 17% respectively with a standard deviation of 4%.

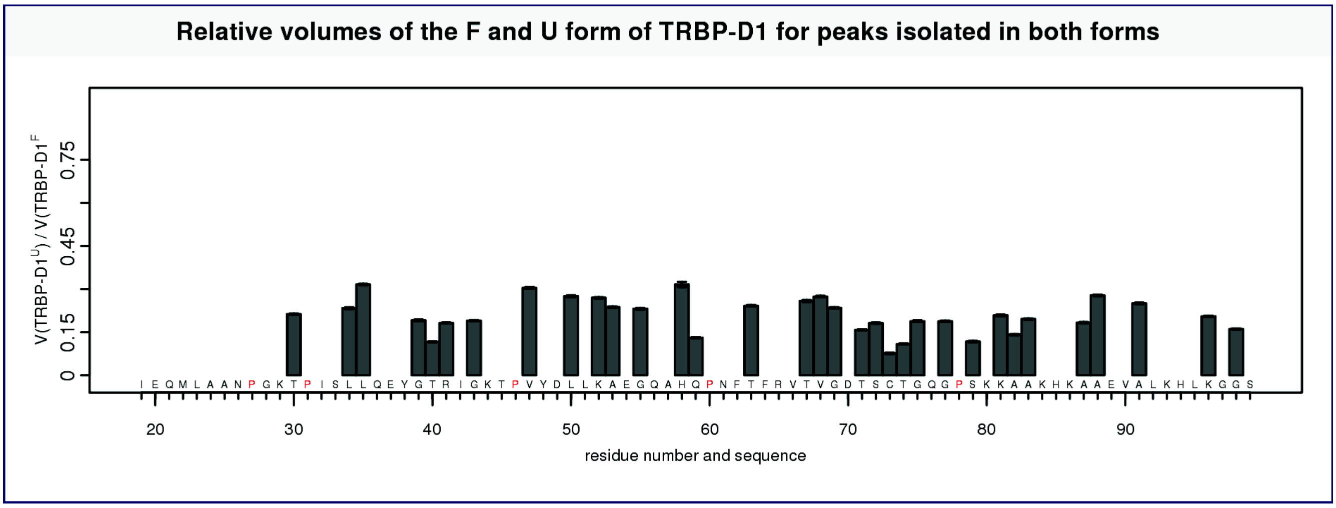


Figure 15.19: Relative volumes of the peaks of TRBP-D1^{U} over TRBP-D1^{F} for the well resolved peaks in the HN spectrum of both forms. The experiment used is an unsaturated Het-NOE experiment with no mixing time and a recycling delay of 5 seconds. It was recorded on a 1.4 mM sample of TRBP-D1 at 600 MHz and 25°C. Error bars (small) represent +/- one standard deviation estimated by simulating 100 ratios using the spectral noise and assuming a Gaussian distribution.

15.8.3.4 Attempt to estimate the folding/unfolding kinetics of TRBP-D1

In order to estimate the values of the exchange constants and the ratio between the two forms, a series of EXSY experiments was performed by varying the mixing time 0 ms, 20 ms, 40 ms, 60 ms, 80 ms, 100 ms, 120 ms, 140 ms, 160 ms and 180 ms. The analysis of these spectra was complicated by the peak crowding. No more than 4 peaks had the four peaks fully resolved. Integration of the volume of the peaks and attempts to fit them with the equations given in **table 12.2** showed that some curves that could be fit for a few residues but the moderate signal to noise ratio did not enable to get reasonably accurate values. An attempt to globally fitted the data did not enable to retrieve relevant information because of an important heterogeneity in the volume evolution of the different residues. This could be related to the moderate signal to noise ratio and/or a lack of data points for longer mixing times and/or to a process more complex that a homogeneous two state exchange. The normalized trends of the peak intensities for a few resolved residues are given as illustration in **fig. 15.20**. The data suggest an exchange rate time scale in the range 10-100 ms. Further analysis on this kind on data would be interesting to get more precise kinetic information.

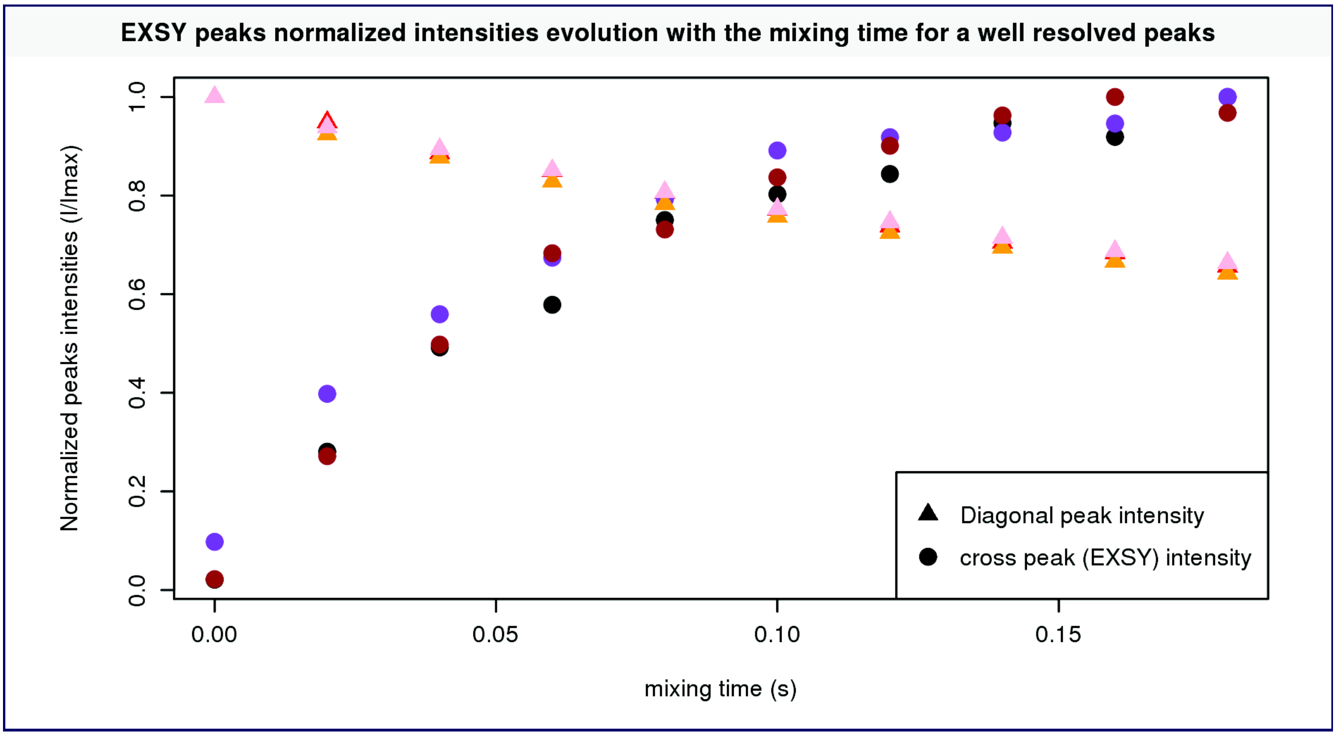


Figure 15.20: Evolution of the diagonal and cross peaks intensities recorded in a series a BEST-EXSY experiments with various mixing times (0 ms, 20 ms, 40 ms, 60 ms, 80 ms, 100 ms, 120 ms, 14 0ms, 160 ms and 180 ms). To show the global trends, the values from different resolved peaks are normalized so that the highest value is 1.

15.8.4 Factors influencing the folding-unfolding equilibrium in TRBP-D1

The goal of this part is to assess whether the second form or TRBP-D1 can be considered as a relevant phenomenon *in vitro* or just a bias linked to the physico-chemical environment used.

15.8.4.1 Effect of environmental conditions

In this part some the potential effects of the physico-chemical environment used in the biophysical studies on the formation of the second form is analyzed.

Molecular denaturants The exchange between a folded and an unfolded form can be obtained experimentally by adding denaturant molecules (urea, guanidine hydrochloride, sodium dodecyl sulfate ...) to a protein solution or with some special protein in a system with extremes (too low or too high) salt amounts. The phosphate buffer used through the study has an ionic strength of 150 mM, which is very close to the mean physiological conditions in humans so the second possibility seems unlikely. Although not known as a denaturant, the only molecule that could affect the folding of the protein in the purification is the Triton-X100 detergent used during the cell lysis. Extensive washes, many dialysis and 2 column purification are used after the first NiNTA step with Triton-X100 so together with the existence of the full folding/unfolding equilibrium there should not be any Triton at the end of the purification. So the folding/unfolding phenomenon detected does not relate to the presence of a denaturing molecule in solution. Extreme pH and temperature (too high or too low) are also known to induce the folding and unfolding of proteins.

pH effects The pH of 6.5 used in this study is close but does not exactly match the physiological one of about 7. To analyze the effect of the pH on the folding/unfolding equilibrium, a protein sample of TRBP2-D1 has been produced with the standard protocol and split into three subsequently dialyzed in parallel at 4°C during 16h in the NMR buffer at three different pH: 5.5, 6.5 and 7.5. The resulting HSQC spectra recorded at 25°C are plotted in **fig. 15.21**. Peaks of the U form are detected in the three pH conditions (6.5 - 7.5). An increase of the volume of the peaks of the U form is observed as the pH decreases. The same experimental parameters, including the short recovery delay (0.2 s), has been used for the three BEST-HSQC experiments. This delay is not optimal for a precise quantification of the population of the two states U and F. Peak volumes were quantified for a few intense and well resolved glycine peaks with detectable intensities for the U form at the 3 pH. For the last two reasons, the ratios for these selected peaks are not good estimate of the mean for the whole sets of peaks, but represents the trend. The present study at pH 6.5 compared to pH around 7 displace the folding-unfolding equilibrium towards the unfolded enough to allow an easy detection and study by NMR.

The overlay of the HSQC also shows that there is a region of the protein, the β1-β2 / N terminus of the β2 sheet that undergoes the largest chemical shift changes upon the pH change. This region is composed of residues 56-`QAHQPNF`-62. This effect very likely corresponds to the protonation of the only protonable side chain in this loop, the histidine 58. The imidazole side chain has a pKa of about 6 which is in the pH region explored here. Being part of a flexible loop according to the NOE data of TRBP-D1, solvent accessibility of this histidine is thus expected.

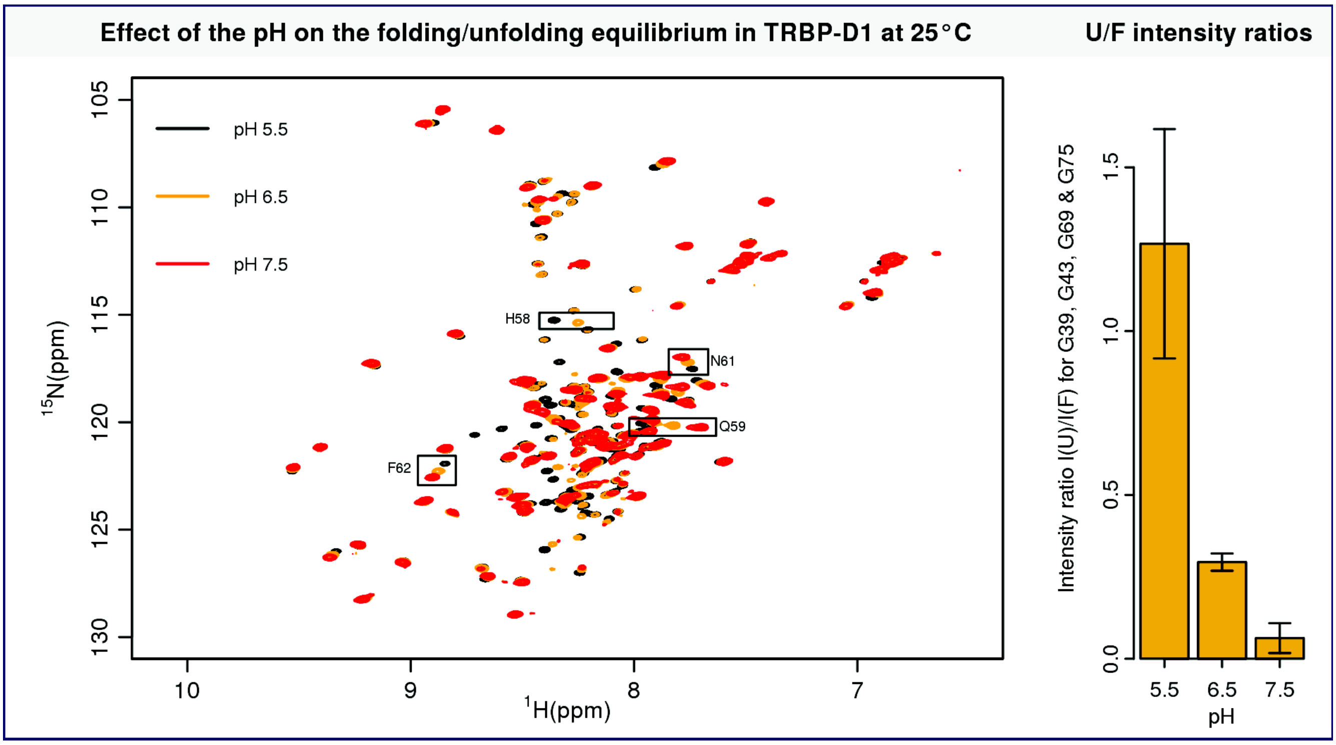
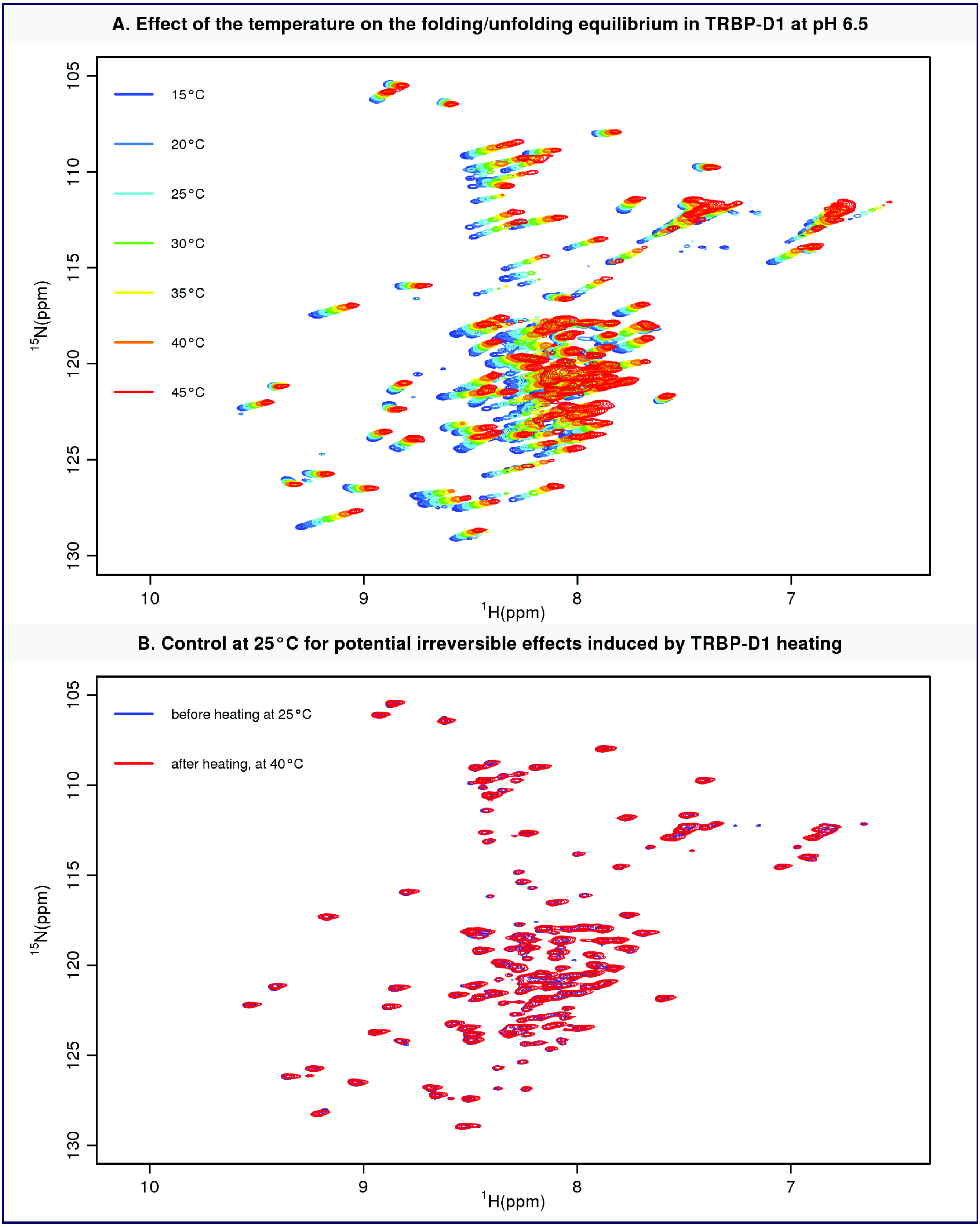


Figure 15.21: Effect of the pH on the folding/unfolding equilibrium in TRBP-D1 at 25°C. **A.** Overlay of three BEST-HSQC spectra recorded on TRBP-D1 at pH 5.5, 6.5 and 7.5 with the same acquisition parameters. Peaks of sequential residues undergoing the larger chemical shift changes present in the β1-β2 loops are surrounded by rectangles and labeled. **B.** Comparison of the ratio of peaks intensities for a few resolved and intense glycine peaks detectable in the U and F state at the three pH. The error bars indicate +/- one standard deviation.

Temperature effects To analyze the effect of the temperature (25°C in the previous experiments), a series of BEST-HSQC spectra of a sample at pH 6.5 has been recorded at different temperatures. In order to avoid potential slight inter sample variability, only one sample was used for the different measurement. Since temperature increase can alter some proteins, the spectra were recorded by increasing the temperature, from 15°C to 45°C by 5°C increments. An additional last spectrum was recorded at 25°C to check whether some alteration of the sample occurred during the heating experiment. Importantly, the experiment is performed in 23.5 mM phosphate buffer which second pKa (7.2) is only very weakly dependent on the temperature ($\frac{dpKa_2}{dT} = -0.028$ at 100 mM) so the effect of pH is neglegible and the experiment essentially reports the temperature effects. The results of this experiment are presented in **fig. 15.22**.



A. Effect of the temperature on the folding/unfolding equilibrium in TRBP-D1 at pH 6.5

B. Control at 25°C for potential irreversible effects induced by TRBP-D1 heating

Figure 15.22: Effets of the temperature on TRBP-D1. **A.** BEST-HSQC experiments of TRBP-D1 from 15°C to 45°C by 5°C increments. **B.** Control experiment. A spectra of the protein sample before the experiment at 25°C is compared to the spectra at 25°C of the sample that underwent the heating experiment.

Importantly, on **fig. 15.22** B., no changes are detected between the spectrum of the in relation to the heating experiment including no increase of the population of the U state. The peaks volumes of the state TRBP-D1^{U} increases with the temperature whereas TRBP-D1^{F} decreases. This is likely a change of the population of the two states. The largest chemical shifts changes are observed for the U state which is perfectly consistent with the fact it is an unfolded form so highly sensitive to the environment. The absence of changes in the control experiment imply that the folding/unfolding evolution as the temperature change is reversible, which is consistent with the previous conclusion of the existence of the equilibrium.

Folding-unfolding exchange at physiological pH and temperature conditions Eventually in order to get closer to physiological conditions, a spectrum has been recorded at pH 7, 37°C **fig. 15.23** from a sample at pH 6.5 dialyzed at pH 7. We see that peaks of the second form are present. Despite the signal to noise of the spectra recorded is rather low, its shows the presence of some peaks that by comparison to the spectrum at 25°C and pH 6.5 correspond unambiguously to the U state of the protein. Additional peaks are visible at the center of the spectrum but are not identifiable by direct spectrum comparison. Identifiable peaks correspond to residues distributed homogeneously along the sequence. These observations confirm the existence of the folding/unfolding equilibrium at a temperature and pH closer to the physiological conditions in human cells. These conditions are however defavorable for a structural analysis by NMR because of a lower protein stability and because of peaks of the main F form present at much lower intensity that at pH 6.5, probably because of a faster exchange of the amide proton with the solvents (about 80 times faster at pH 7 compared to 6.5 (29)). The observations made in this paragraph underline the interest to study the protein samples in the conditions used here (pH 6.5, 25°C), making both the folded state and folding/unfolding event more detectable by NMR and not too far from the physiological ones.

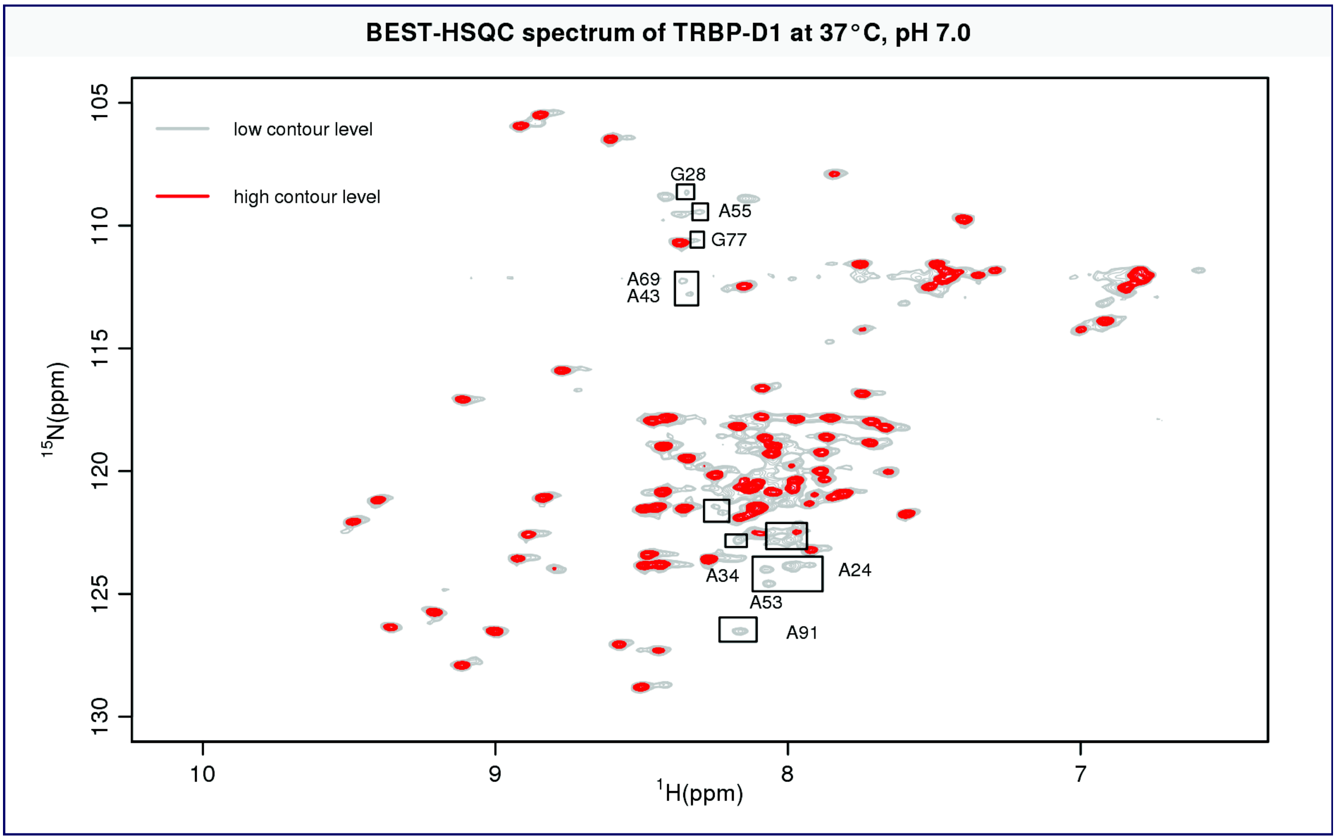


Figure 15.23: BEST-HSQC spectrum of TRBP-D1 in recorded at pH 7.0 and 37°C at 800 MHz in [23.5 mM potassium phosphate, 100 mM KCl, 10 mM $MgCl_2$ and 5 mM β-mercaptoethanol, 5mM TCEP]. The rectangles indicate a few peaks of the unfolded state (U) of TRBP-D1 identified by comparison with the assigned spectrum at pH 6.5 and 25°. Unambiguous assignment for a few residues of the U state is also given.

15.8.4.2 Analysis of the effect of the construct boundaries

The aim of this part is to study whether the folding/unfolding phenomenon of TRBP2-D1 could be the result of the truncation effects. Indeed since TRBP-D1 construct it is only a part of TRBP, the presence of the rest of the C-terminal polypeptide could stabilize it; for instance if a secondary structure motif has been cut in the construction design or if C-terminus folds with the first domain. In this case, the folding-unfolding process would be irrelevant and would be a truncation artifact.

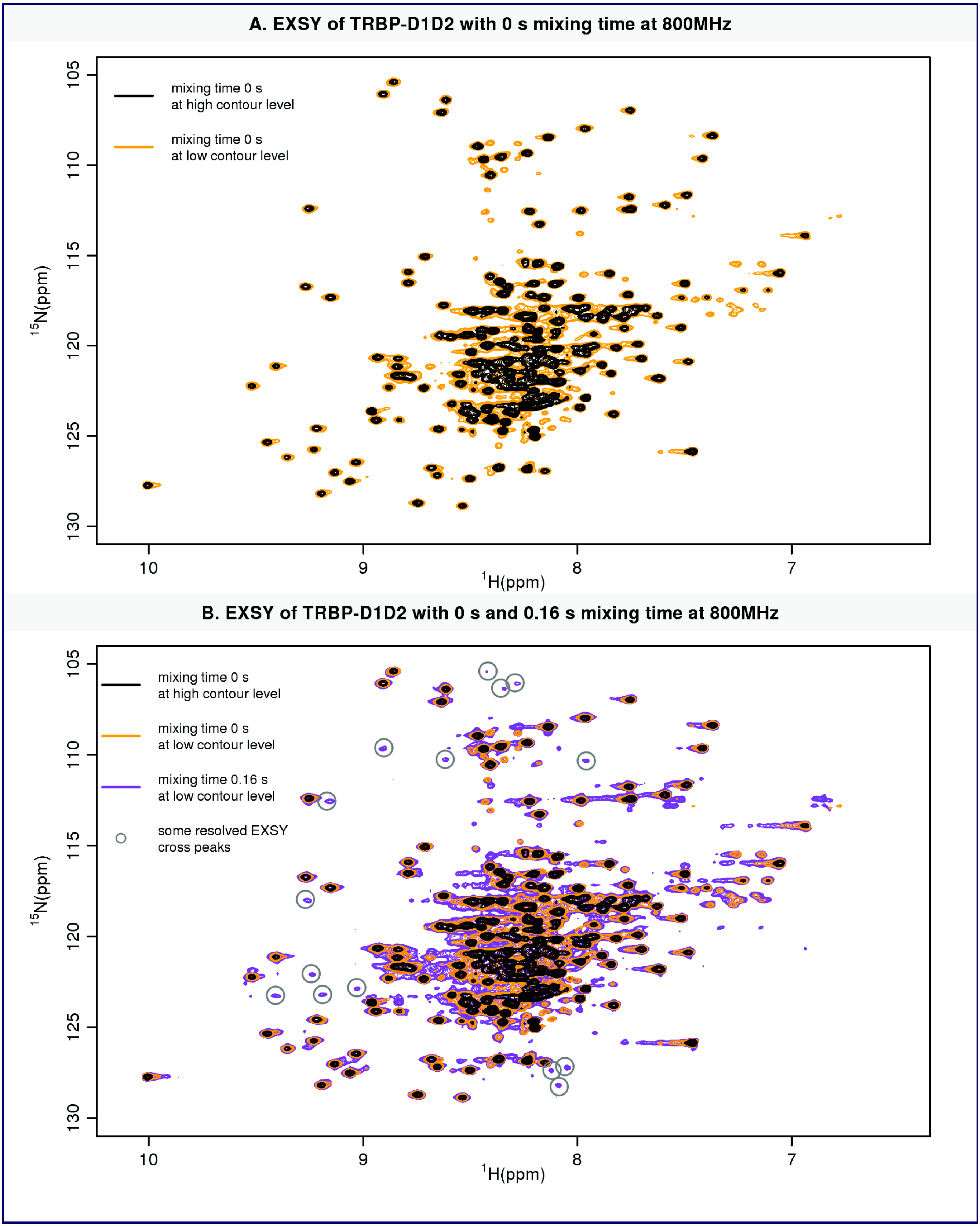


Figure 15.24: BEST-EXSY experiments of TRBP-D1D2 at 25°C with mixing times of 0 s and 0.16 s recorded on a sample of 480 μM at 800 MHz in [23.5 mM potassium phosphate pH 6.5, 100 mM KCl, 10 mM $MgCl_2$ and 5 mM β-mercaptoethanol, 5mM TCEP]. **A.** Spectra with 0 s mixing time. To easily see the main species of TRBP-D1 compared to the other species, the spectrum is plotted at high and low contour level in black and orange respectively. **B.** Spectra with 0.16 s mixing time, overlaid with the spectra A. Some well resolved EXSY cross peaks are highlighted with gray circles.

Folding/unfolding on the first domain in the presence on the linker and second domain of TRBP The peaks of the second form of TRBP2-D1 are also detected in the TRBP-D1D2 construction and the exchange happen at similar time scale, according to one EXSY experiment performed with a mixing time of 160 ms on an 800 MHz spectrometer, with a protein concentration of 480 μM (**fig. 15.24**).

Folding/unfolding on the first domain of TRBP and TRBP isoforms The first isoform of TRBP (TRBP1) has an extended N-terminal region that could also stabilize the first domain. TRBP-D1 and TRBP-D1D2 correspond to the second isoform of TRBP (TRBP2) with 21 N terminal residues in the N-terminus than the first isoform of TRBP (TRBP1). In TRBP-D1 and TRBP-D1D2 3 additional N-terminal residues of the first isoform and a residual glycine from the TEV cleavage site are present so there is not a truncation at this extremity. However an effect on the first 20 residues of TRBP1 on the folding/unfolding process could help to understand the biological selection of the two isoforms. The folding/unfolding equilibrium could also occur in both isoforms.

To clarify this NMR experiments on TRBP1-D1 were performed **fig. 15.25** . The HSQC and EXSY experiment show that the peaks characteristic of the U state are also present in the TRBP1-D1 construct and with the two forms in exchange ; the relative intensities of the U and F state in the two form recorded in similar conditions are not distinguishable. Thus, similarly to the extension of the C-terminus, the presence of the exchanging unfolded form of the first domain of TRBP is not influenced by the length of the N-terminus. Although a more quantitative study could have been performed on the TRBP1 isoform, the highly similar HSQC spectra and the absence of *in vivo* data suggesting a relevant behavior difference between the two isoforms has not motivated it.

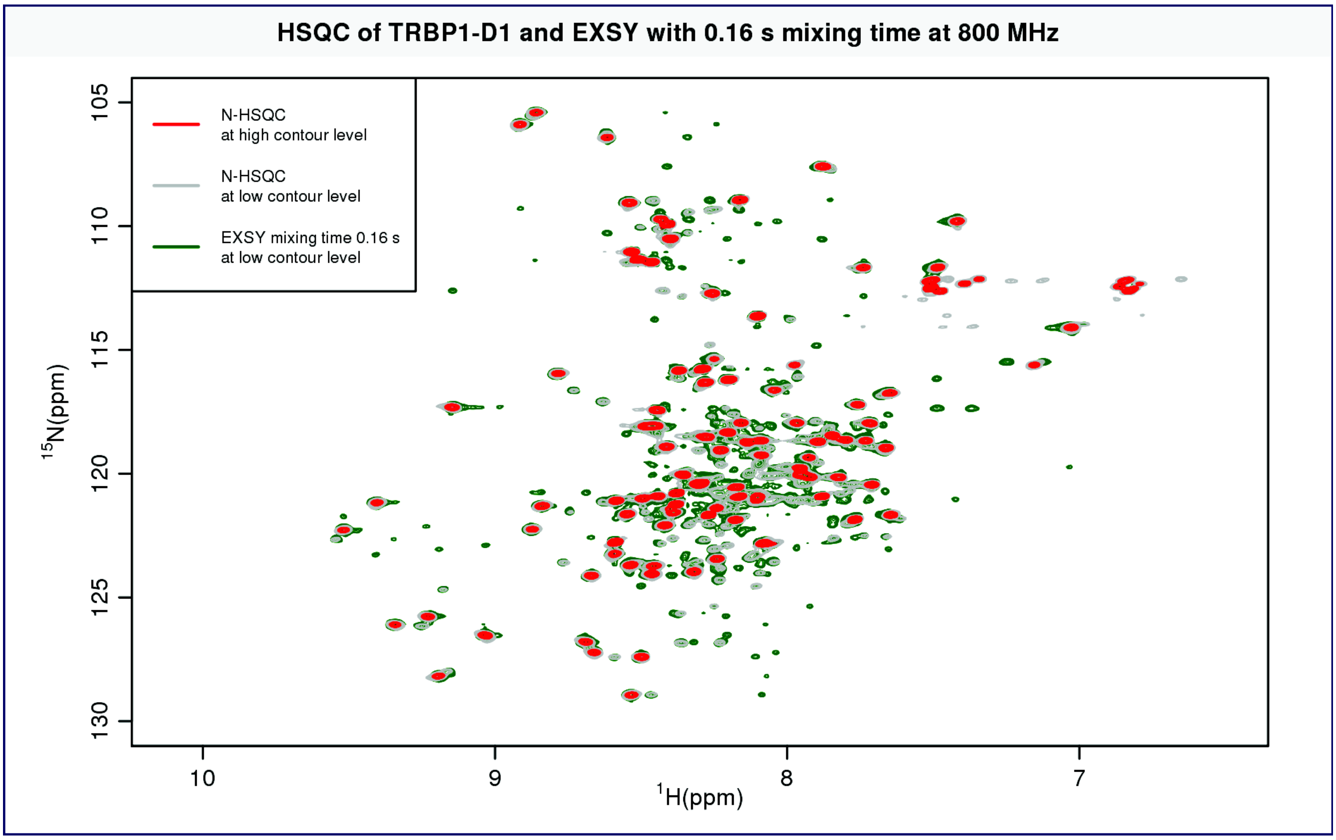


Figure 15.25: Evidence of an exchange in TRBP1-D1 (first dsRBD of TRBP with the 22 residue N-terminal extension) at 250 μM in phosphate buffer at pH 6.5, 25°C. **A.** N-HSQC and a BEST-EXSY experiment (mixing time 0.16s) recorded at 800 MHz are overlaid.

15.8.4.3 Analysis of the buffer effect on the folding/unfolding equilibrium

Finally, to test a potential effect of the phosphate ions on the observed folding-unfolding process, some samples were dialyzed in 50 mM PIPES buffer at pH 6.5 with 100 mM KCl and 5 mM TCEP. Both the folded and unfolded form were detected (HSQC and EXSY), with the same characteristics than in the phospahate buffer used **fig. 15.26**. So the use of a phosphate buffer is not responsible to the phenomenon observed.

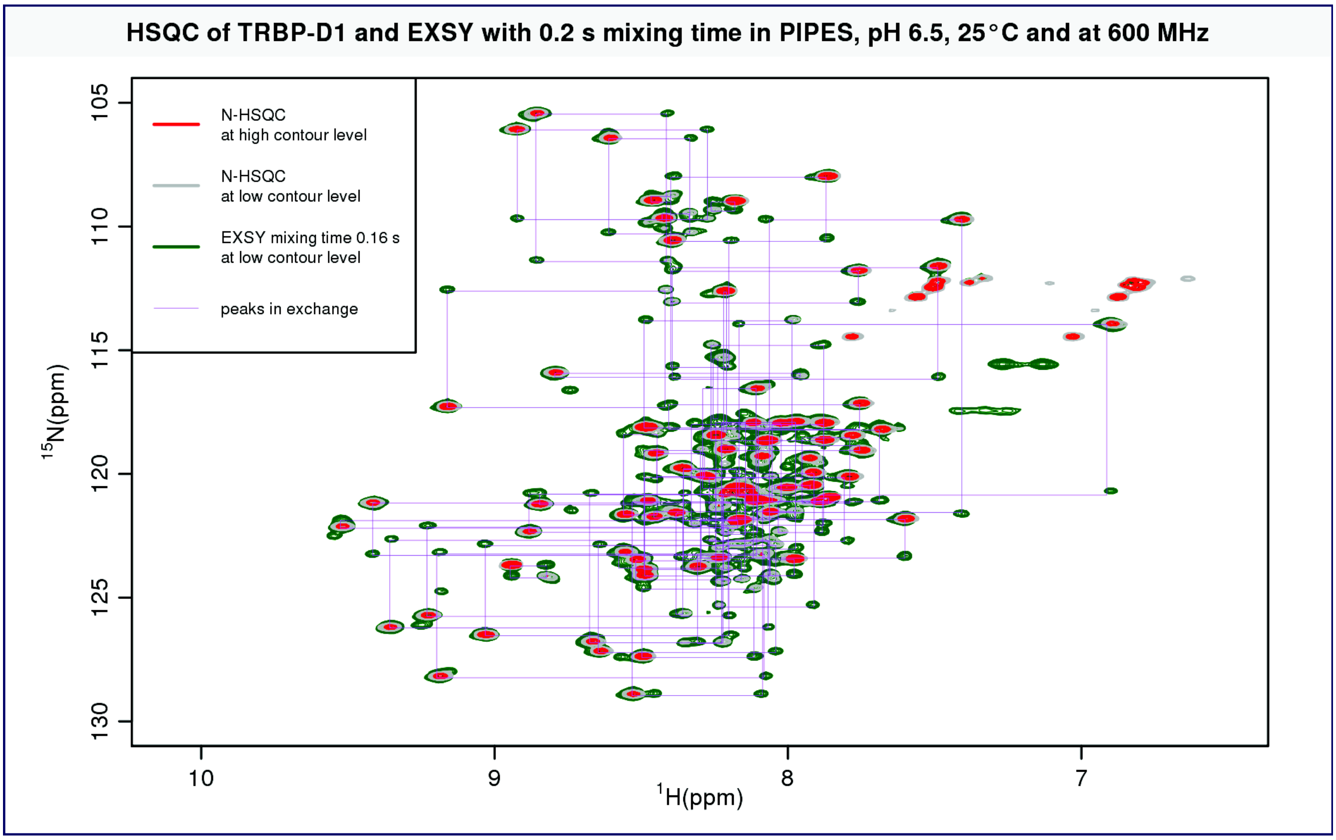


Figure 15.26: Evidence of an exchange in TRBP-D1 at 750 μM in PIPES buffer at pH 6.5, 25°C with 100 mM KCl and 5 mM TCEP. An N-HSQC and a BEST-EXSY experiment recorded at 600 MHz are overlaid. The assignment of TRBP-D1^{F} and TRBP-D1^{U} at 25°C in the phosphate buffer used in this study was overlaid to show the network of peaks in exchange.

15.8.4.4 Conclusion

As the result of these analysis, the existence of the newly detected folding-unfolding equilibrium of the first domain of TRBP appears to be an intrinsic property of this protein *in vitro*, not caused by obviously adverse physico-chemical conditions in the media, nor by truncation effects and is not isoform dependent.

15.8.5 Hypotheses to explain the origin of the folding/unfolding equilibrium of the first domain of TRBP

It is hard to get mechanistic information on the folding/unfolding processes using biophysical techniques recording properties of an ensemble of molecules, even for site resolved solution NMR experiments. Fast pulsing techniques and tricks to slow down the folding process have been applied to favorable systems. Ultra-high time-resolved site specific information would be necessary. The goal of this paragraph is just to present a few hypotheses that could explain the special behavior of the first dsRBD of TRBP using the data from this study and the ones reported in the literature. Since the folding/unfolding equilibrium was not detected in the second dsRBD (but could exists at different time scale and/or with much lower U population), a comparison of the sequence and structures of these dsRBDs can be interesting. The structures of the folded form of D1 has been obtained by X-ray diffraction (PDBid: 3LLH) whereas the one of the second domain was solved by NMR (PDBid: 2CPN).

15.8.5.1 Sequence comparisons

Sequences differences As the first two dsRBDs of TRBP only share 34% identity, the search of sequence determinants for the different behavior of TRBP-D1 and TRBP-D2 is not straightforward (**fig. 3.2**). On the structural point of view, according to the published structural data, the main form of the first two TRBP domains have a very similar fold but a fold comparison might not be relevant as the structures were solved in different conditions, in solution for TRBP-dsRBD2 and in a crystal for TRBP-dsRBD1. A few sequence features are discussed here.
- A tryptophan appears inserted at the position 14 of the alignment (152 in TRBP) in the α1-β1 loop of the second domain of TRBP. This residue is one of main the difference between the fold of the first two domains of TRBP as underlined by the study that reported both structures (290). The aromatic ring of the tryptophan interacts with the side chain of lysine 12 (150 in TRBP) through a cation-π interaction. This study has found using circular dichroism that the thermal stability of the first domain of TRBP is significantly lower than the one of the second domain. The authors used a protein construct very similar to the ones reported in this work but didn't removed the purification tags in N-terminus. Their result get more sense in light of the folding/unfolding equilibrium of TRBP-D1 discovered in this thesis and thus both studies appear consistent. To understand the thermodynamical differences, they suggested that the cation-π interaction present in the second domain but not in the first one stabilizes the second domain. They found that the mutation of the tryptophan residue involved in that interaction to an alanine make the second domain of TRBP unfolded. However this result is not a definitive proof of a role for the cation-π in the stability of TRBP-dsRBD2. Another mutation at another place could have led to the same result. Even if the tryptophan is actually an important residue for the second domain the reason for the destabilization of the construct might not be caused just by its removal.
- The alignment shows that 3 prolines are present at almost the same position in the each dsRBD, but the first domain of TRBP has an additional proline at the N terminus of the helix α2, at the center of a long region that appear rather rigid in the NOE time scale and that seem important for the dsRBD fold as mentioned in the first part. Proline cis/trans isomerization is a common source of kinetic step limiting process in the folding of proteins and this reaction in this region could contribute to the phenomenon observed.

Anyhow, the origin of the folding/unfolding equilibrium in TRBP-D1 is potentially not localized to a few residues and could be a global result related to the overal chemical composition of the dsRBD. The combination of the special features at the C and N terminus of the two helices (no stabilizing tryptophan, and presence of a proline respectively) with the shorter C terminal helix could cooperate in the low domain stability.

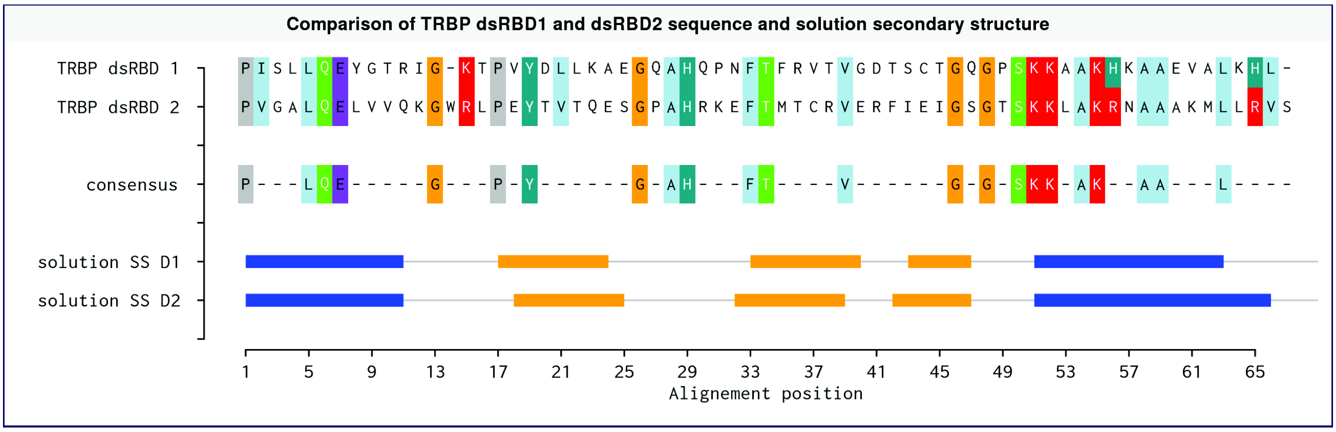


Figure 15.27: Sequences and secondary structure comparison between TRBP-D1, TRBP-D2. **A.** Multiple alignments was performed with the MUSCLE program (65). The Shannon's information theoretic entropy (246) was used with a 10 letter alphabet to measure the residue diversity ([V,I,L,M], [F,W,Y], [S,T], [N,Q], [H,K,R], [D,E], [A,G], [P], [C], [-]) and normalized, implemented by the R bio3D package. The color scheme is the following one: (A, V, M, I, L, F, C and W) are colored in light blue, (S, T, N and Q) are colored in green, D,E and R,K are colored in purple and red respectively, H and Y in blue-green and G are colored in orange. Aligned residues with a normalized entropy of conservation value below 0.40 are not colored. The secondary structures of TRBP-D1 and TRBP-D2 are the ones obtained in this work whereas for PACT-dsRBD1, they are derived from its reported solution structure (PDBid: 2dix) with Stride algorithm (89).

15.8.5.2 Towards an understanding of the folding/unfolding process of TRBP-D1

I propose here potential research directions that could contribute to understand the determinants of the folding-unfolding event in the first binding domain of TRBP :
- *Mutational studies* Providing that the phenomenon observed could be a general property of some dsRBDs, a good method to understand the origin of the folding-unfolding process could be to generate mutants that destabilize the second domain so that a folding unfolding phenomenon appear *and* other mutants that stabilize the first domain to diminish the unfolding propensity. However as the potential observed effects are not obviously related to the changes we expect to induce, a correct analysis of these multiple mutants would require to analyze the mutants by NMR and to solve their solution structure, which would have been very time consuming especially with part of the signal coming from the unfolded forms that would give a lot of "contaminating" peaks in the NOESY spectra. Even the assignment of the HN backbone resonances have been strongly complicated by the presence of the minor form that also gave

strong 3D signals, often overlaid with peaks of the main form. Mutations to understand folding processes have often been performed on proline motivated by the idea that the cis/trans isomerization can impact the folding kinetics. Such a study would require many mutations per site on various sites to be conclusive as the result cannot just be interpreted as a proline effect: the effect of the residue used to replace the proline can also have an important effect on what will be observed. A good example of these ambiguities has been published (26) for the protein CRABPI, a protein characterized by a slow folding phase. This work shows that mutating proline 85 into an alanine abolish the slow folding phase whereas a mutation to valine does not change the folding kinetics. Some mutation of the tryptophan 152 of TRBP-dsRBD1, aimed to get insight into the difference of thermal stability between the first two domains of TRBP have been published during the course of my PhD (290) as already mentioned but they did not fulfill all the criteria required to provided a definitive explanation.

- *Cis-trans proline isomerization contribution analysis* A simpler approach could easily assess whether the proline isomerization have a contribution to the folding-unfolding exchange detected. Prolyl isomerase could be added to an NMR sample and HSQC and EXSY experiments could monitor whether or not the equilibrium has shifted towards the folded form. Even if this isomerization appears to be responsible for the folding-unfolding process, the equilibrium would still be relevant *in vitro* since in the absence of the isomerase, the protein should go back to the folding-unfolding equilibrium.

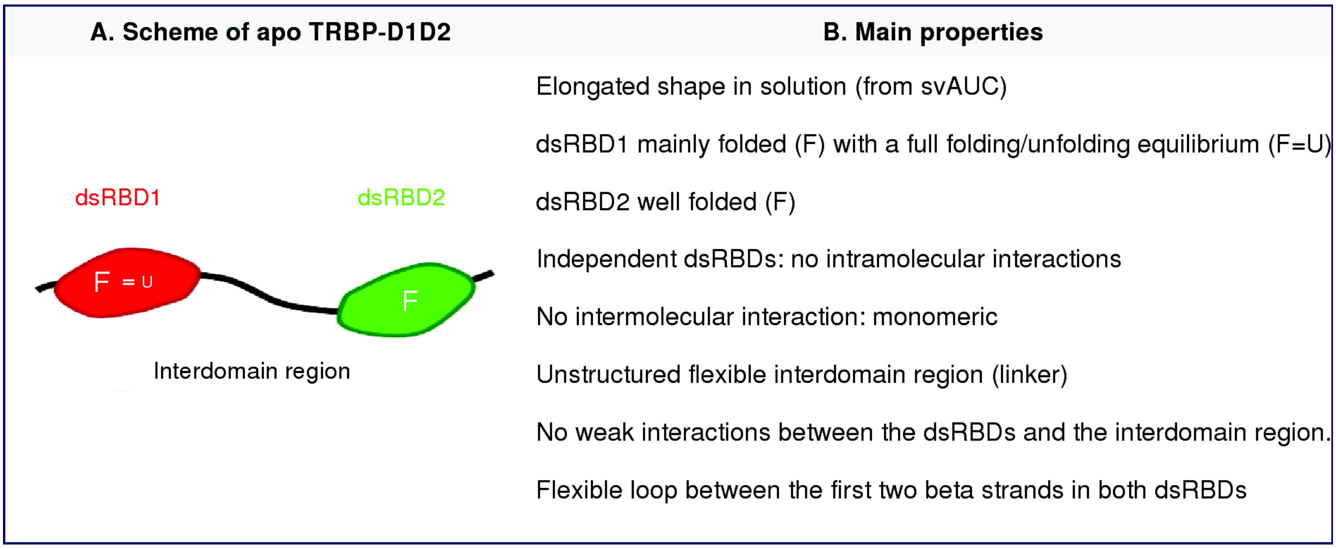
15.8.6 Perspectives regarding the folding/unfolding of the first domain of TRBP

The first remark that is worth saying, as already underlined, is that the solution structure behavior of a protein can be different from the one obtained by the analysis of the x-ray diffraction pattern of protein crystal. Indeed the "x-ray" structure of TRBP-D1 shows a folded protein and it is very likely that the crystallization has shifted the folding/unfolding equilibrium toward the folded form. The possibility to study the solution structural and dynamical behavior of a macromolecule is a major advantage of NMR spectroscopy over crystallography.

The second remark is more a semantic one. In the last 10 years it has been shown that a large portion of the human proteome is composed of unfolded protein or contain unfolded region that are now called intrinsically disordered proteins (IDP). NMR spectroscopy has enabled to get many insights into these proteins showing, among other things, that these proteins are often not completely random coils. Using this terminology, the first domain of TRBP-D1 can be considered as an IDP with a high propensity of structure. To my knowledge the first domain of TRBP is the first reported evidence of a full folding / unfolding equilibrium in a human protein domain at pH and temperature around physiological conditions. Thus this example underlines well that there are not two worlds of proteins, the folded ones and the intrinsically disordered ones but a continuum, the two latter cases being extremes situations. The TRBP-D1 protein appears as a special member of this continuum and it would be in the future to see whether similar behavior are found in other human proteins and could be related to any / many function(s). This requires a careful analysis of the NMR spectra to understand all the peaks and minor forms present in them.

This original TRBP-D1 behavior among the dsRBDs family appears a unique property apart from the previous reported behavior of some SH3 protein domains that surprisingly resemble a lot what has been detected here. SH3 domains are a class of domains found in many proteins participating to signaling pathways. Like dsRBDs, SH3 domains have also many structural variations. Despite most SH3 structures have been reported to form stable structures in solution, a group has shown the existence in *Drosophila* of an equilibrium between a folded and an unfolded state in the N terminal SH3 domain at pH 6 of an adaptor proteins composed of an SH2 domain surrounded by two SH3 domains, called drk (302; 301; 76). Another group reported the same phenomenon in the mammalian homolog GRB2 but at pH 3.5 (102). The case of drk has been extensively studied and is characterized by two almost equally populated states with a mean rate of exchange of 0.9 s (76) with no evidences of an intermediate state in the folding pathway (200). A mutation study has been performed for the drk N terminal SH3 domain where a conserved glycine was occupied by a threonine (200) the threonine to glycine mutation stabilized the domains.

15.8.7 Summary of the characterization of the apo-RNA binding region of TRBP



Elongated shape in solution (from svAUC)

dsRBD1 mainly folded (F) with a full folding/unfolding equilibrium (F=U)

dsRBD2 well folded (F)

Independent dsRBDs: no intramolecular interactions

No intermolecular interaction: monomeric

Unstructured flexible interdomain region (linker)

No weak interactions between the dsRBDs and the interdomain region.

Flexible loop between the first two beta strands in both dsRBDs

Figure 15.28: Summary of the main properties of apo TRBP-D1D2 found in the present study.

The main properties found here in TRBP-D1D2 are summarized in **fig. 15.28**. These properties will constitute a basis for the study of the interaction between this region of TRBP and miR-155 precursors.

Chapter 16

Characterization of apo pre-miR155 and miR155/155* duplex

16.1 Homogeneity analysis of microRNA 155 precursors

As the RNAs were produced *in vitro* and purified on gel to be sure to get a high purity, the technique to purify and analyze the purity are similar. An example of a urea-PAGE of pre-miR-155 and miR-155/miR-155* and a SEC chromatogram of a pre-miR-155 are given in **fig. 16.1** and **fig. 16.2**.

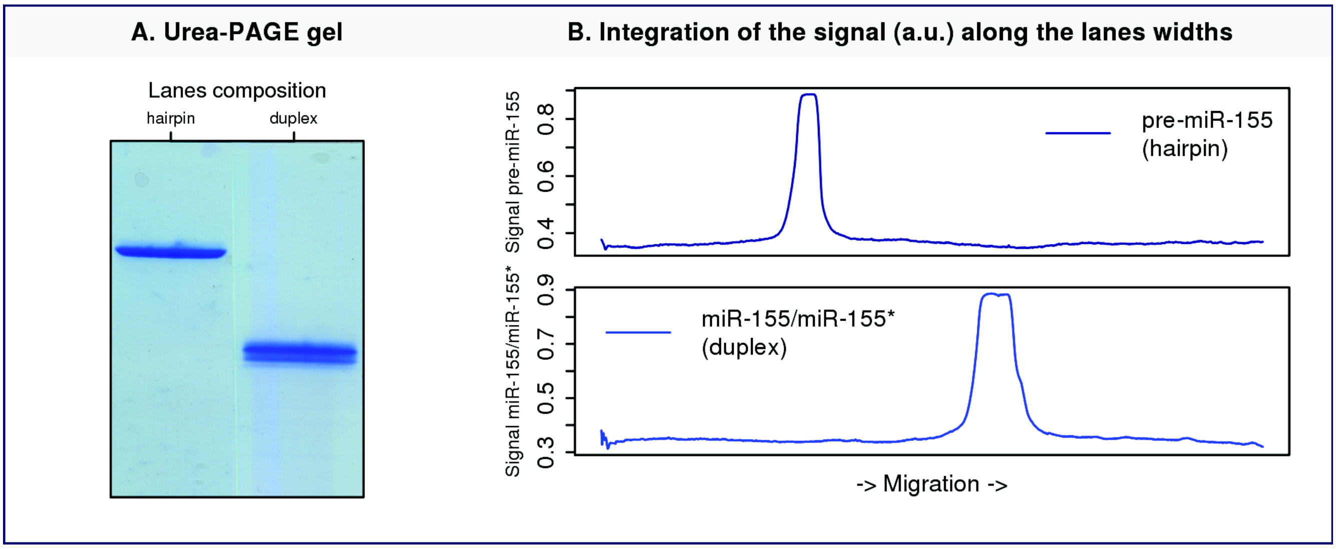


Figure 16.1: Urea-PAGE gel analysis with 15% acrylamide of pre-miR-155 (hairpin) and miR-155/miR-155* (duplex) constructs after their purification and activation. Methylene blue was used for the staining. **A.** Two lanes juxtaposed, extracted from a larger gel. **B.** Integration of the gel lanes with arbitrary signal units. The signal saturates around 0.9.

One the gel, **fig. 16.1**, one band is obtained as expected for the hairpin. Two bands are resolved in the lane of miR-155/miR-155 which corresponds to the 23 bp miR-155 and 22 bp miR155*. The purity of the samples was considered sufficient to perform biophysical studies with them.

16.2 Oligomerization analysis of pre-miR-155 by SEC-MALLS

SEC-MALLS experiment Because of the low yields of miR-155/miR-155*, most biophysical analysis were performed with pre-miR-155; the duplex was mainly used in the interaction study by NMR. A SEC analysis with mass estimation with an on-line MALLS detector was performed for pre-miR-155. The protocol is described in **chapter 8**. A sample of 50 μL of pre-miR-155 at 200 μM (4 mg/ml) was loaded on the SEC column. Because of the very large absorption of the RNA in the UV range, this sample saturated the detector so the resulting chromatogram is given in **fig. 16.2** using the refractive index detector signal.

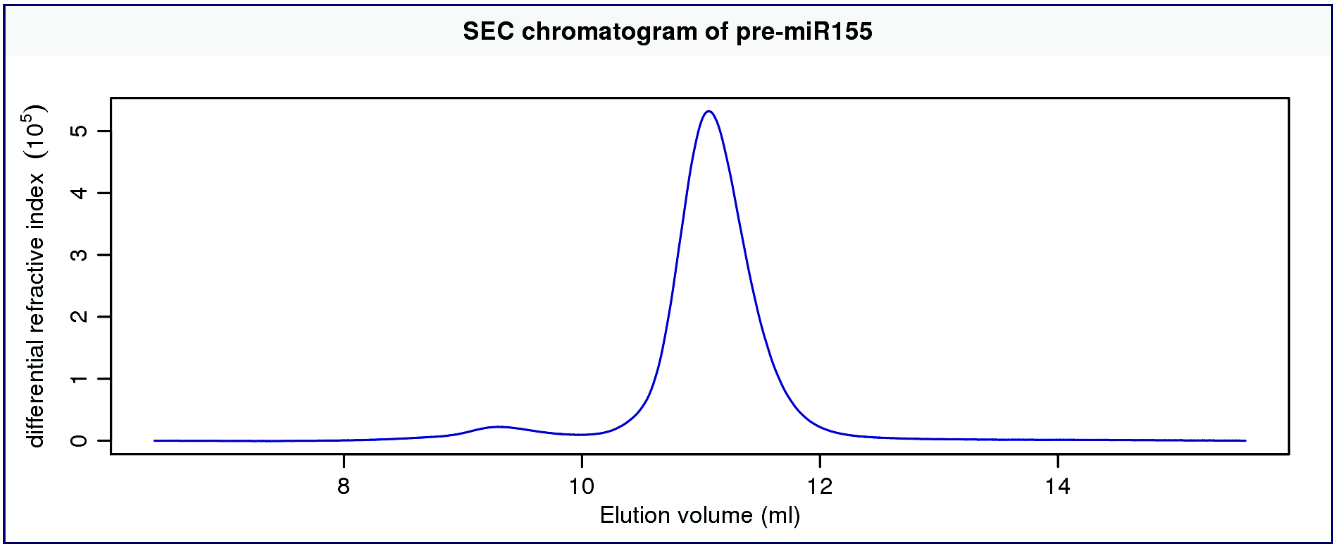


Figure 16.2: SEC refractive index profile of a sample at 200 μM of pre-miR155 in phosphate buffer loaded on an analytical Superdex S75 10/300 GL column (GE Healthcare) coupled with an in-line refractive index detector (RI 2000, Spares) used to follow the differential refractive index of the SEC eluate relative to the solvent (1.34)

SEC-MALLS data analysis Two peaks are present in the chromatogram, despite the SDS-PAGE gel of pre-miR-155 is characterized by only one band. This indicates the presence of two species of pre155 in the sample. The MALLS molar mass estimation of the molecule in the main peak matches very well what is expected for monomeric RNA (19.6 kDa): 19.14 kDa with standard deviation (sd) of 0.76 kDa. The amount of the species corresponding to second peaks is too low for an accurate MALLS estimation. Two hypotheses can explain the presence of these two species of RNA :

- Hairpin interaction. Because of the high density of negative charge in the RNA and providing the secondary structure prediction is valid, it is hard to conceive the possibility for this RNA to form dimers of hairpins. Such interactions have already been observed for perfect duplex of DNA with blunt ends. In such case, staking interactions can lead to the dimerization of part of the sample. However, no such extremities are present in the RNA construct. This hypothesis seems very unlikely.

- Partial duplex formation. The RNA have been folded using a quick cooling of a hot (96°C) solution of RNA. This should favor the kinetic product of the reaction, which is the intramolecular folding leading to the hairpin. Another fold are also possible (although not predicted by the mfold algorithm (309)), corresponding for instance to an imperfect two strands duplex that contains two times more base pairing than the hairpin fold **fig. 16.3** Presence of this product could be favored under thermodynamic control (slow folding, low temperature). Such hairpin and homo-duplex coexistence has already been observed, for instance for a 79 nucleotide RNA (258).

A. **mfold predicted secondary structure of pre-miR-155**

```
             c   a        u-   cc
  ggaaugcuaau gug uagggguu  uug  \
acccuuacgauua-uac-auccucag  aac  /
                          uc   cu
```

B. **A possible duplex secondary structure**

```
             c   a        u-   ccuc   cu
  ggaaugcuaau gug uagggguu  uug    caa  gacuccua-cau-auuagcauuccca
acccuuacgauua-uac-auccucag  aac    guu  uuggggau gug uaaucguaagg
                          uc   cucc   -u        a   c
```

Figure 16.3: **A.** Secondary structure of pre-miR-155 according to mfold prediction. **B.** Author proposed secondary structure to account for the detection of a small fraction of dimerized pre-miR-155.

To validate MALLS estimation, this result has to be compared with the assumption used to derive the molar mass values :

- $\mathcal{A}_{MALLS}1$. The estimated masses very close to the monomeric mass, correspond to RNA chains that are well below the laser wavelength (690 nm) so this assumption is fulfilled.
- $\mathcal{A}_{MALLS}2, 3$. The assumptions 2 and 3 are hardly testable and their potential effects have been discussed in the methodological part.
- $\mathcal{A}_{MALLS}4$. Providing a intermolecular base pairing formation is the mechanism that generates the species with a large molar mass, two stables RNA species are present in solution without transient interaction. This is required for the system to get close to the assumption 4.

16.3 Oligomerization analysis of pre155 by svAUC

svAUC experiment In order to analyze the pre-miR-155 sample with a different theoretical framework to confirm the SEC-MALLS observations, a svAUC analysis has been performed. ^{1}H 1D NMR experiments have shown that the RNA samples were stable for at least 5 months. They were however used much before this long delay. This stability enables to study the RNA using a svAUC experiment (about 10 hour). The svAUC analysis was performed on two samples of 1 μM and 2 μM in the using the protocol presented in the methodological part. In particular the experiment is performed in the same phosphate buffer than in the others biophysical studies. The low concentrations of RNA enable to avoid saturation of the UV detector and to get closer the important assumption in svAUC requiring to work at very low concentrations.

svAUC results The concentration profiles recorded during the experiment at 280 nm are given in **fig. 16.4**. The data are rather noisy. This could come from impurities on the surface of the cells windows (*e.g.* oil) or damaged cells windows.

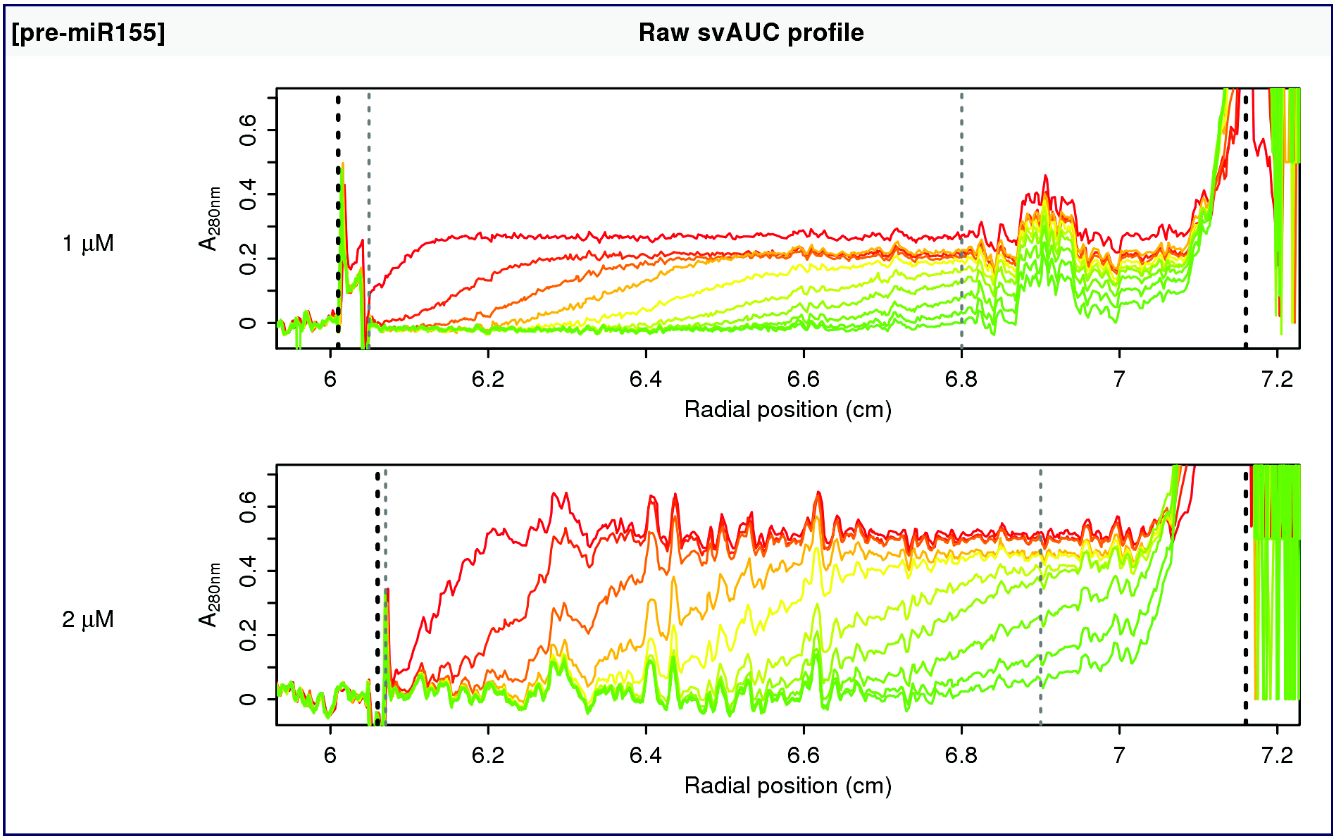


Figure 16.4: Superimposition of the 10 first scans recorded at 280 nm in the svAUC experiments for pre-miR155 at 1 μM and 2 μM. The bold dashed lines indicate from left to right the position of the liquid meniscus and the position of the cell bottom. The two lighter lines in gray indicate the zone used to fit the data. The right light gray dashed line indicate an approximate position after which the effects of the concentration of the sample at the cell bottom became obvious in the sedimentation profiles

svAUC data analysis The result of the diffusion deconvoluted $c(s)$ model fitting is given in **fig. 16.5**.

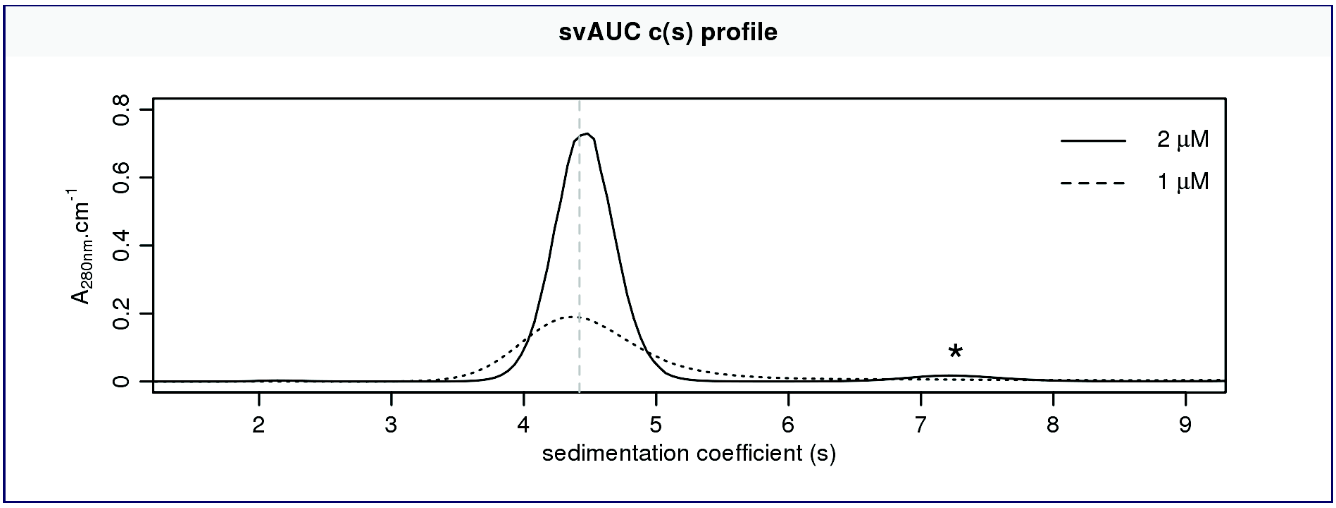


Figure 16.5: Superimposition of the diffusion deconvoluted $c(s)$ profiles of pre-miR155 at 1 μM and 2 μM respectively. The dashed gray line is positioned at the mean of the maximum of both distributions. The star (*) indicates a minor species, with represents 4% of the main signal species sedimenting faster.

Because of the rather noisy svAUC data for the precursor and the reduced range of concentration tested, it is not possible to assign an origin to the slight change of sedimentation coefficient at 1μM and 2 μM around 4.4 S. However, the c(s) profile of the sample at 2 μM shows the presence of a minor larger species sedimenting at 7.2 S. This signal from this species corresponds to 4 % of the main one and is consistent with the small fraction of larger RNA species detected previously in SEC-MALLS experiments. Assuming a concentration independency of s_{exp} for the RNA, a modeling with a single sedimenting species can be applied. This model, without assumptions about the shape of the sedimenting species leads to a molar mass estimation of 19.4 kDa which is close to the real mass of 19.6 kDa. Taking into account the assumptions used in the derivation of the model used, this estimation is considered very good. As a result, this svAUC analysis confirms the mainly monomeric behavior of pre-miR155, with a minor part of the population existing as a larger species. Additionally, fixing the mass in the $c(s)$ model to fit the data leads to a frictional ratio of 1.4. This value is larger than the one expected for a spherical particle (1.25) and indicates an elongated particle. This is expected for an hairpin structure. Using this frictional ratio it is possible to predict the sedimentation coefficient of an RNA dimer with a similar frictional ratio (6.75 S), which can be adjusted to 7.2 with a frictional ratio of 1.31. Thus the minor peak in svAUC is consistent with the dimerization of a small fraction ($\approx$ 10%) of pre-miR-155. Because of the moderate yield of pre-miR-155, the low proportion of dimers, the handling risk associated with additional purification steps to try removing the potential dimer and the very high probability that the larger RNA species is in slow equilibrium with the hairpin, it has been decided to keep the RNA sample as it is, without additional purification steps.

16.4 Secondary structure of miR 155 precursors - NMR assignment of imino protons

16.4.1 Secondary structure of pre-miR-155

Imino protons of guanine and uracil bases protected from exchange with the solvent i.e. involved in hydrogen bounding (93) have NMR signal in the 10-15 ppm range (93). The imino spectra of pre-miR-155 are characterized by about 20 resolved peaks that show existence of base paring in this RNA **fig. 16.6**.. All RNA samples samples produced were systematically analyzed by 1D (1H) NMR spectroscopy to check for a correct *in vitro* folding and they shown identical spectrum to the one reported in **fig. 16.6**. In order to study experimentally the validity of the secondary structure predicted for the pre-miR-155, assignment of the imino spectra of this RNA was performed. For unlabeled RNA, imino proton of assignment is commonly performed by ^{1}H-^{1}H-NOESY experiment. Lowering the temperature (e.g. 5°C) increases the stability of the base pairing so it usually enables to detect the less stable base pairs at 25°C, especially the ones present at the boundaries of hybridized regions. In order the check the relevancy of the RNA assignment at 5°C for an inference of its structure at 25°C (the temperature at which all interaction studies were performed), the effect of the temperature on the 1D ^{1}H imino spectrum on pre-miR-155 was analyzed. **fig. 16.6**.

The temperature titration shows only moderate shifts and intensity changes mainly related to the separation of two peaks at 14 and 12 ppm. These moderate changes suggest that the base pairing pattern of the RNA in the range [5°C-25°C] does not change importantly apart from expected stabilization of weak base pairs. It is thus meaningful to study the secondary structure at 5°C for an assignment at 25°C. The spectra used for the assignment of pre-miR-155 are given in **fig. 16.7** and assigned imino protons are represented in **fig. 16.7**. Two independent paths are detected in the HH-NOESY exactly as predicted by the theoretical secondary structure. Additional peaks detected in the 1D imino proton spectrum and in the HH-NOESY indicate than additional weaker base paring interactions exist which is also consistent with the predicted secondary structure. In particular a peak at 11.2 very likely correspond to the two G:U in position 24. Dots have been used to represent these weak interactions in the structure of pre-miR-155 in **fig. 16.7**.

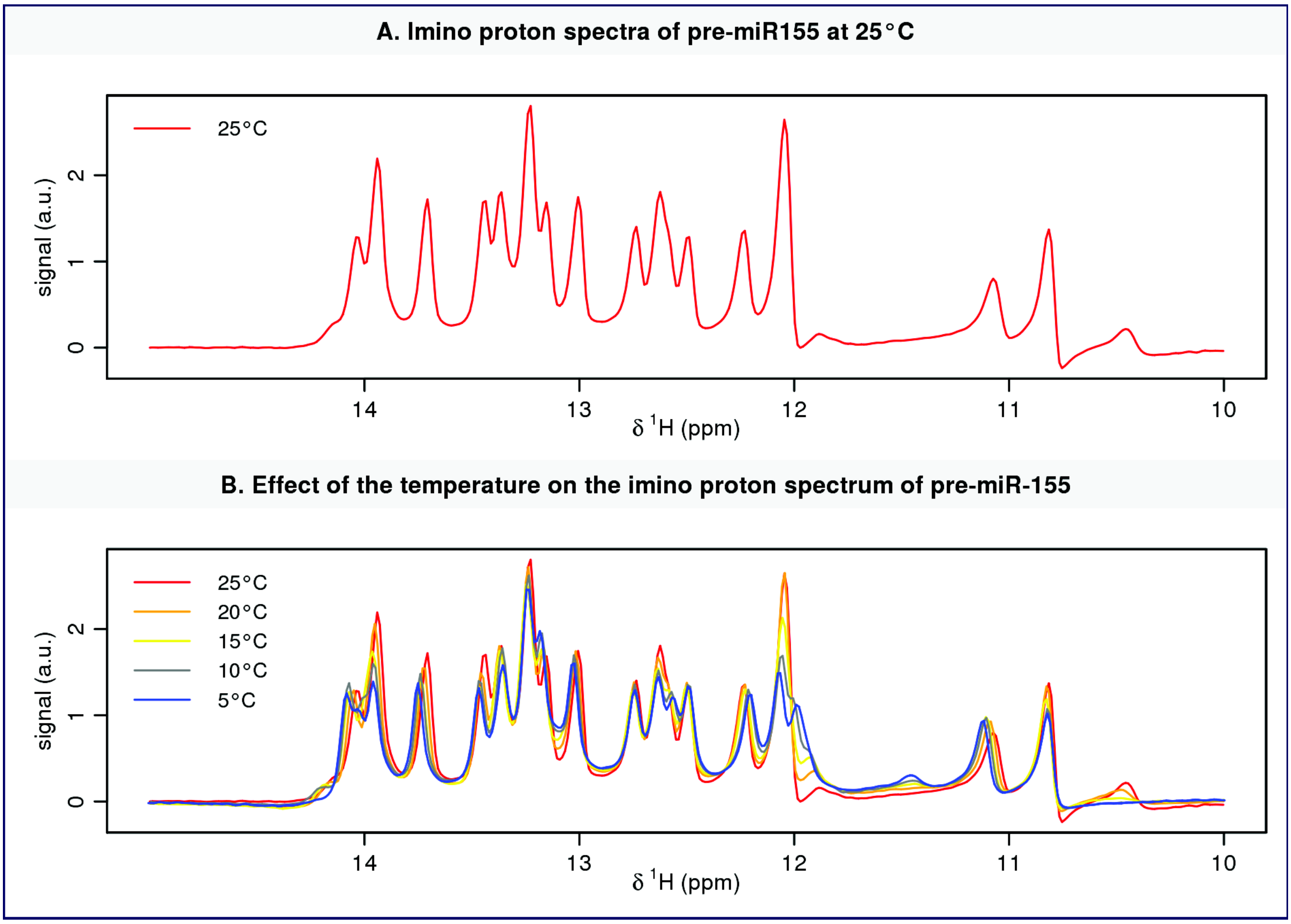


Figure 16.6: **A.** Imino proton spectrum of pre-miR-155 of a 890 μM sample recorded at 25°C at 800MHz. **B.** Effect of the temperature in the range [5°C, 25°C] on the imino proton spectrum of pre-miR-155. In order to facilitate the comparison of the spectra, they were scaled so that the sums of the intensities are equal.

16.4.2 Secondary structure of miR-155/miR-155*

Direct comparison between the assignment imino proton spectrum of pre-miR-155 with miR-155/miR-155* suggests that the first base pairing region (nt 1 to 10 and 50 to 60) has a similar base paring structure in miR-155/miR-155*. It is not possible to make conclusion for the second segment of base pairs (nt 17 to 22 and 41 to 46) because of important shift of the peaks compared to pre-miR-155. This is expected as this second region is shorter in the duplex compared to the hairpin and since its chemical composition is different: G:U and G:C at position 21 and 22 respectively are replaced by two C:G. The absence of the strong signal from the assigned G:U pair at position 21 in pre-miR-155 is thus consistent with the composition of the duplex. In addition to the peaks of the first base paring region, four imino peaks are detected and imply a base pairing in the second region, so the base paring structure proposed in fig. 16.8. There is an excellent match of the imino protons peaks expected to be found in the longest helix of the duplex precursor and an expected less correspondance for the shorter helix due to the two mutations and the absence of the helix.

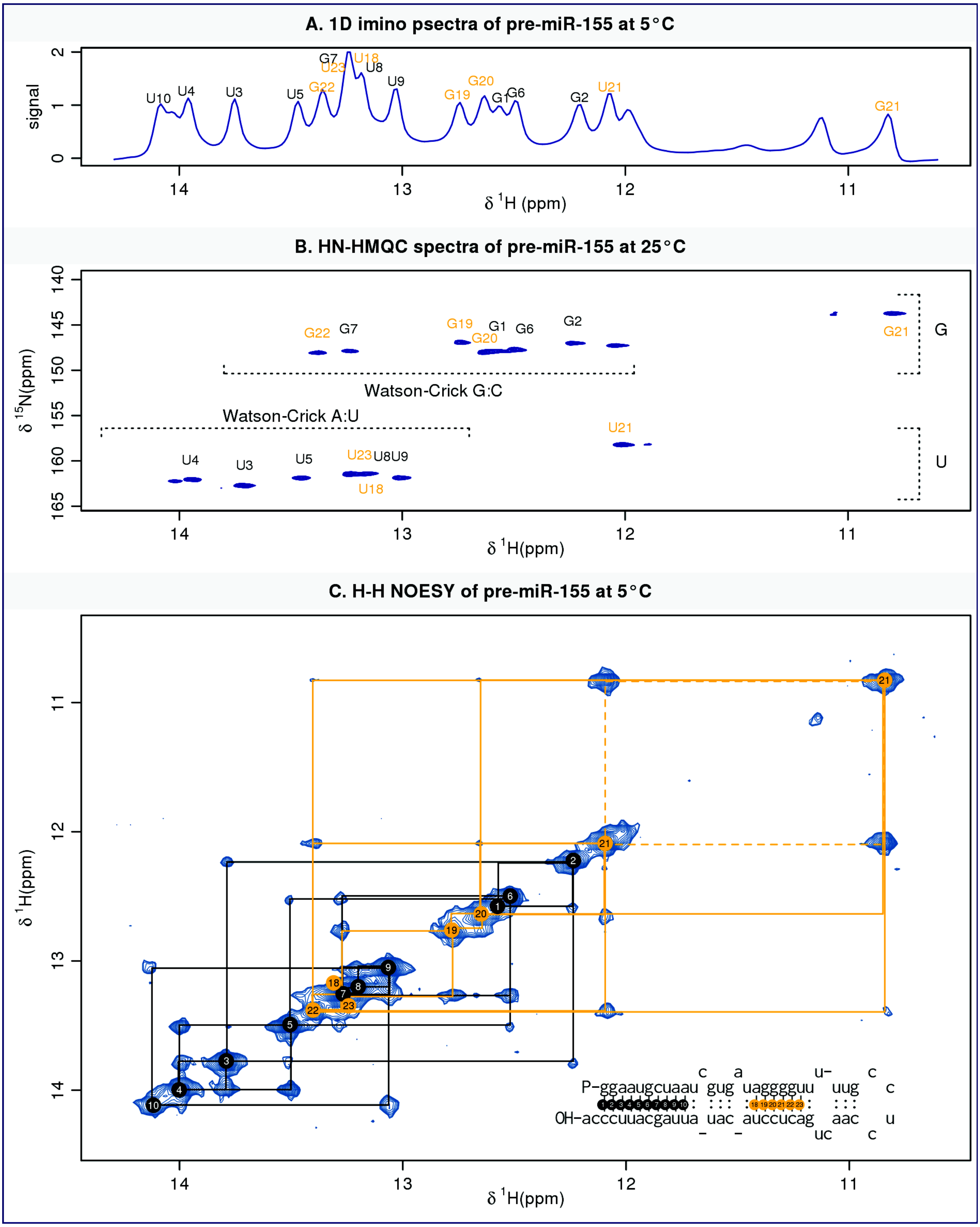
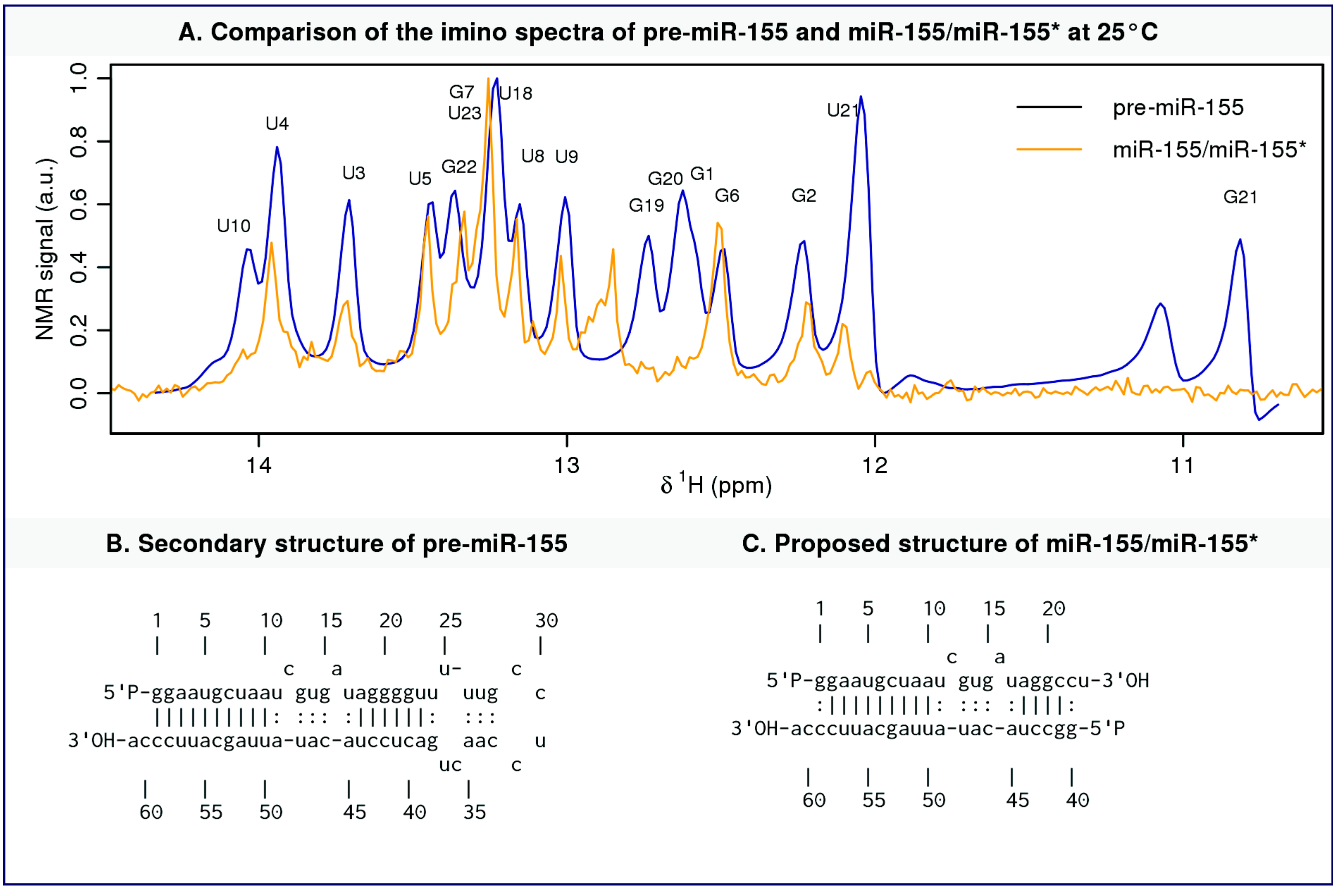


Figure 16.7: Spectra used to assign the imino protons of pre-miR-155 (engineered). **A.** imino proton spectrum of pre-miR-155 at 890 μM recorded at 800 MHz and 5°C in the phosphate buffer with 5% $2H_20$. **B.** SOFAST NH-HMQC of pre-miR-155 at 890 μM recorded at 600 MHz and 25°C in the phosphate buffer with 5% $2H_20$. **C.** HH-NOESY of pre-miR-155 at 890 μM recorded at 800 MHz and 5°C in the phosphate buffer with 5% $2H_20$



B. Secondary structure of pre-miR-155

```
       1   5    10   15   20   25      30
       |   |    |    |    |    |       |
                  c   a        u-      c
   5'P-ggaaugcuaau gug uagggguu  uug   c
       ||||||||||: ::: :||||||:  :::
3'OH-acccuuacgauua-uac-auccucag  aac   u
                               uc     c
      |    |    |      |    |    |
      60   55   50     45   40   35
```

C. Proposed structure of miR-155/miR-155*

```
       1   5    10   15   20
       |   |    |    |    |
                  c   a
   5'P-ggaaugcuaau gug uaggccu-3'OH
       :|||||||||: ::: :||||:
3'OH-acccuuacgauua-uac-auccgg-5'P

      |    |    |      |    |
      60   55   50     45   40
```

Figure 16.8: Comparison of the imino proton spectrum of engineered pre-miR-155 and miR-155/miR-155*. **A.** Overlay of the imino proton spectrum pre-miR-155 and miR-155/miR-155* recorded at 25°C. The assigned imino protons of pre-miR-155 are indicated. **B.** Secondary structure for pre-miR-155 according to the HN-HMQC and HH-NOESY study. **C.** Proposition of secondary structure for miR-155/miR-155* according a the proton imino spectrum comparison with assigned pre-miR-155 (**A.**). The base numbering is the one used for pre-miR-155 to facilitate comparison.

16.5 Summary of pre-miR-155 and miR-155/miR-155*

The different biophysical methods applied to pre-miR-155 have shown that this RNA is mainly monomeric with a small fraction (< 4%) present probably as a dimer in solution. The monomer adopts a hairpin structure in solution as expected and has an elongated shape in solution. Presence of base pairing in the miR-155/miR-155 duplex is also consistent with the expected structure.

From these informations on miR-155 precursors and on TRBP-D1D2, TRBP-D1 and TRBP-D2 in their apo form, a biophysical study of the interaction between the two kinds of macromolecules could be performed.

Part V

Characterization TRBP-D1D2 interaction with miR-155 precursors

Goals In the case of TRBP interaction with microRNA precursor, as underlined in the bibliography, the stoichiometry and macroscopic binding parameters have not been studied in depth in the literature. In particular it was still unclear whether there is one RNA binding to each dsRBD, one or more protein interacting per RNA or a mixture of these cases. Another key question was to determine whether the RNA binding region of TRBP interacts differently with the hairpin or duplex type of microRNA precursors. The latter question is equivalent to know the potential role of the loop on the binding parameters or to compare the role of the RNA binding region of TRBP before and after Dicer cleavage for miR-155 precursors system. To address the last point, the magnitude of the binding constants of the microRNA binding region had also to be estimated for both types of precursors. Then deeper study of the interaction at atomic level, using atomic sites specific information by NMR, will be analyzed.

As previously explained in part II, the combined used of ITC and svAUC can enable to characterize the macroscopic equilibrium between interacting species. Then NMR was used to get site specific informations :

- **Chapter 17**: Study of TRBP interaction with miR-155 precursors by ITC
- **Chapter 18**: Stoichiometry analysis of the RNA binding region of TRBP by svAUC
- **Chapter 19**: Site specific characterization of the RNA binding region of TRBP with microRNA precusors : two independent binding domains
- **Chapter 20**: Conclusion and discussion of the main results

Chapter 17

Study of TRBP interaction with miR-155 precursors by ITC

17.1 ITC experimental design and implementation

The molecular ratio of TRBP and microRNA in their complex was not known at the time of the design of these experiments. There is the possibility that this ratio might not be unique: an excess of protein or an excess of RNA can potentially lead to different complexes. To avoid missing one of them, ITC experiments were performed in two ways, either by adding more and more RNA to a low concentrated protein sample, either by doing the opposite. Because of the methodological consideration on the concentration of the species in the calorimeters, this resulted in experiments either with the protein in the syringe and the RNA in the cell or the other way. The concentrations used in these experiments were mainly determined by the available material and the necessity to perform repetition of the experiments. The different experiment conditions are summarized in **tab. 17.1**. ITC experiments were conducted using a MicroCal ITC-200 at 25°C using the steps detailed in the methodological part II, Chapter 10.

17.2 ITC analysis of TRBP-D1D2 interaction with miR-155 precursors

17.2.1 ITC datasets

Different typical ITC datasets for the interaction between TRBP-D1D2 interaction with pre-miR-155 and miR-155/miR-155* duplex as well as dilution controls are given in **fig. 17.1**. In each case the titrations generated a binding isotherm characterized by a sigmoidal shape. The control experiments showed only very minor power change during the injection. Instead of estimating these constant negligible effects with a risk to add errors to the main experiments, the

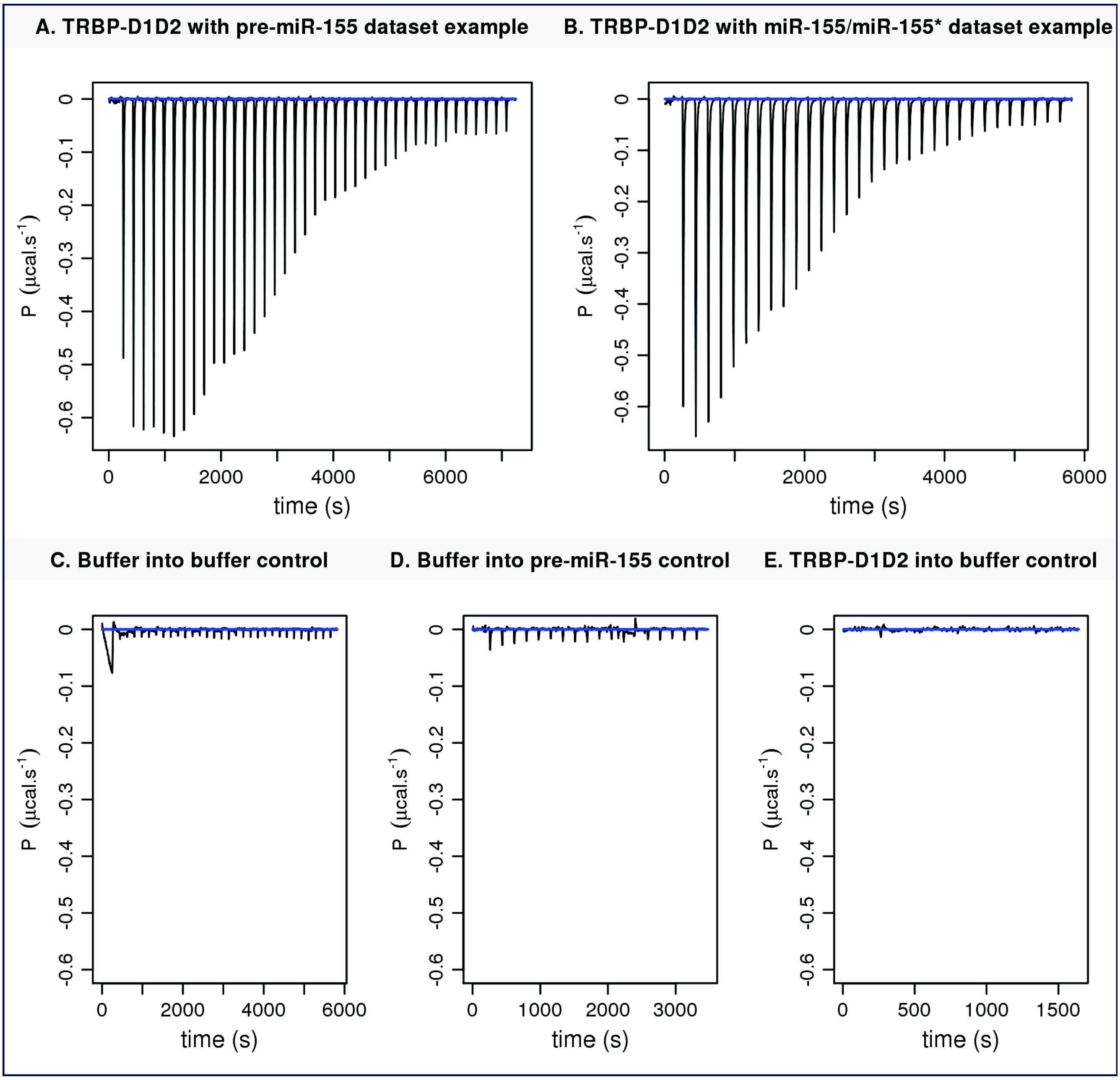
Table 17.1: Concentration of protein and RNA used in the different ITC experiments. The component with the higher concentration is systematically the one in the syringe. The first line full of zeros indicate multiple control experiments with buffer in both the syringe and the cell.

TRBP-D1 (μM)	**TRBP-D2** (μM)	**TRBP-D1D2** (μM)	**pre-miR-155** (μM)	**miR-155/miR-155*** (μM)
0	0	0	0	0
0	400	0	0	0
0	0	250	0	0
0	0	0	300	0
0	0	59	300	0
0	0	500	30	0
0	0	42	275	0
0	0	43	275	0
0	0	50	300	0
0	0	54	300	0
0	0	250	40	0
0	0	450	30	0
0	0	500	40	0
0	0	500	30	0
0	0	375	0	33
0	0	400	0	25
45	0	0	300	0
48	0	0	275	0
800	0	0	30	0
800	0	0	30	0
0	20	0	160	0
0	36	0	150	0
0	37	0	300	0
0	1200	0	35	0
0	1200	0	35	0
0	1200	0	50	0

experiment was analyzed without control subtraction but with a constant baseline as described in **section 10.2.1**.

17.2.2 Application of the fitting procedure

The fitting procedure applied is the one described in **section 10.2.5.3**. The relevance of the different ITC data analysis assumptions 10 will be discussed at the end of the analysis. Two kinds of experiments were performed with TRBP-D1D2 a miR-155 precursor, either with TRBP-D1D2 added in an RNA solution either with an RNA solution injected in a TRBP-D1D2 solution. For convenience, from now the experiments with protein added to the RNA will be called "forward" experiments and the opposite titration, "reverse" experiments.


A. TRBP-D1D2 with pre-miR-155 dataset example

B. TRBP-D1D2 with miR-155/miR-155* dataset example

C. Buffer into buffer control

D. Buffer into pre-miR-155 control

E. TRBP-D1D2 into buffer control

Figure 17.1: Examples of ITC datasets of TRBP-D1D2 injected into miR-155 precursors recorded on ITC-200 (MicroCal) at 25°C. **A.** Baseline corrected ITC dataset of 500 μM TRBP-D1D2 injected into a solution of pre-miR-155 at 30 μM. 39 injections of 1 μl lasting 2 s spaced by 3 minutes interval were performed. **B.** Baseline corrected ITC dataset of 400 μM TRBP-D1D2 injected into a solution of miR-155/miR-155* at 25 μM. 31 injections of 1.25 μl lasting 2.5 s spaced by 3 minutes interval were performed. **C.** Control experiment with phosphate buffer injected into the same buffer; 31 injections of 1.25 μl lasting 2.5 sec with 180 s time interval were used. **C.** Control experiment with buffer injected into a solution of pre-miR-155 at 30 μM. 18 injections of 4.3 μl spaced by 3 minutes intervals were performed. **E.** Control experiment with TRBP-D1D2 at 250 μM injected into the phosphate buffer. 15 injections of 1 μl lasting 2 s and spaced by 3 minutes.

17.2.2.1 Application of the fitting procedure to forward experiments

Determination of a set of candidate macroscopic models The first step was to determine a set of reasonable macroscopic models that can reasonably fit the data. In particular this aimed to determine whether multiple TRBP-D1D2 can interact with miRNA precursors whether multiple RNA can bind TRBP-D1D2 or whether only one miR-155 precursor interact with TRBP-D1D2. Thus, models to test included {1 protein: 1 RNA} ; {2 protein: 1 RNA} ; {2 RNA: 1 protein}. The folding unfolding equilibrium occurring in TRBP-D1 was not taken into account in the ITC analysis as it would add too many parameters in the models to enable to get meaningful binding and folding parameters. This is further discussed and justified in the light on the NMR study of the interaction in part V. Both models {2 protein: 1 RNA} and {2 RNA: 1 protein} could fit all the forward and reverse experiments with TRBP-D12. The {2 RNA: 1 protein} model could not fit the data at all, meaning that the process could not be explained by this model. This means either that TRBP-D12 do not interacts with two precursors, either that this happen but with a zero enthalpy of reaction. Despite the last possibility seems unlikely, the svAUC analysis is necessary to clarify this ambiguity. To distinguish between the {1 protein: 1 RNA} and {2 protein: 1 RNA} models, forward ITC experiments were performed with as many injections as possible in order to be able to detect potential multiple binding steps. This is illustrated on **fig. 17.2** for two experiments with pre-miR-155.

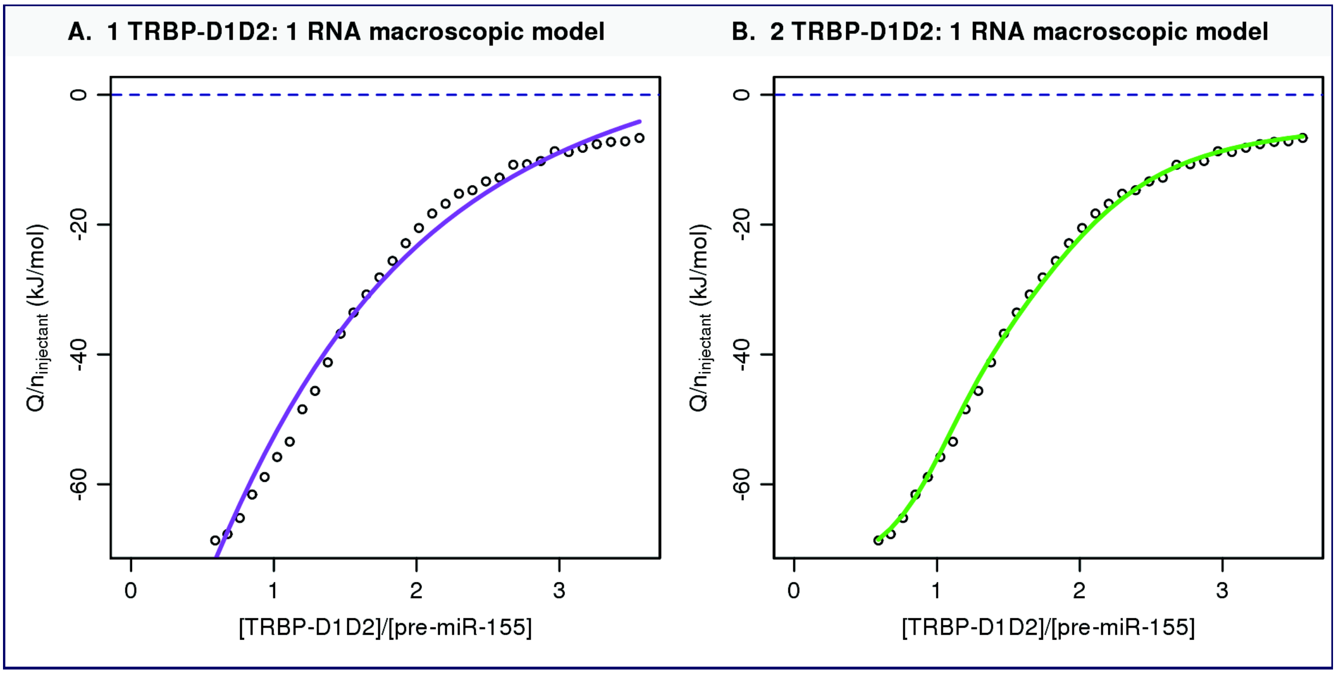


Figure 17.2: Illustration of the fits of the ITC data of the interaction between TRBP-D1D2 and pre-miR-155. The plot represents the integration of the ITC power signal of a baseline corrected dataset corresponding to an experiment with 500 μM TRBP-D1D2 injected into a solution of pre-miR-155 at 30 μM. 39 injections of 1 μl lasting 2 s spaced by 3 minutes interval were performed. **A.** Fit with a 1:1 protein:RNA binding model. **B.** Fit with a 2:1 protein:RNA binding model.

Two experiments with 31 or 39 injections of 1.1 or 1 μl were performed in parallel for both pre-miR-155 and miR-155/miR-155* in order to compare the behavior of these two precursors. These data were globally fitted with both {1 protein: 1 RNA} and {2 protein: 1 RNA} models. The quality of the fit is attested by the low root mean square deviations (RMSD) values in **tab. 17.2**.

Table 17.2: Root mean square deviations and results of the model comparison using Akaike weights for the successive binding models that could be globally fitted to the data for each miR-155 precursors. The stoichiometry column indicate the maximal protein over RNA ratio in the model used

System	stoichiometry	RMSD (kcal/mol)	Akaike weight
TRBP-D1D2 + pre-miR-155	1	0.13	1.10^{-27}
TRBP-D1D2 + pre-miR-155	2	0.053	≈ 1
TRBP-D1D2 + miR-155/miR-155*	1	0.14	7.10^{-31}
TRBP-D1D2 + miR-155/miR-155*	2	0.044	≈ 1

Macroscopic models comparison The comparison of the macroscopic models {1 protein: 1 RNA} and {2 protein: 1 RNA} with Akaike weights in **tab. 17.2** for the globally fitted data show non ambiguously that given the data and these 2 models, the model of successive binding of two proteins on one RNA has a probability of about 1, whereas it is almost 0 for the other model. This proves that the molar ratio that best explain the data is 2 proteins for 1 RNA. Assuming this molar ratio is the actual stoichiometry, which is expected from the behavior of the single partners but require a check with another method (svAUC, Chapter), this means that for both miR-155 precursors, two molecules of TRBP-D1D2 can successively interact with them, leading to 1:1 and then 2:1 protein: RNA complexes. The forward titration appears well suited for the study of such equilibrium, with the complexes progressively formed during the titration.

Binding parameter estimation The selected macroscopic {2 protein: 1 RNA} model enables to get estimates of the macroscopic parameters of the interaction between TRBP-D12 and pre-miR-155 or miR-155/miR-155* (**tab. 17.3**). As explained in **section 6.2.2** the constants estimated are "concentration" dissociation constants and their confusion as real dissociation constant is under the hypothesis of activity coefficients very close to 1, that all the reaction were detected and that the protein:RNA molar ratio correspond to the actual stoichiometry. Anyhow, the estimations show very similar binding parameters for the two RNA precursors. This implies a very similar recognition of these precursors by TRBP-D1D2 with no clear selection effects.

Table 17.3: Result of the estimation of thermodynamic parameters for TRBP-D1D2 with both pre-155 precursors. The standard deviation for each fitted parameter is indicated in parentheses and was determined by running 100 Monte Carlo simulations of each pair of ITC curve followed by a global fitting of each pair of experiments. D1D2: TRBP-D1D2 ; hairpin: pre-miR-155 ; duplex miR-155/miR-155*.

system	Kd_1 (10^{-7})	ΔH_1 (kcal.mol^{-1})	ΔG_1 (kcal.mol^{-1})	ΔS_1 ($\text{Cal.mol}^{-1}\text{K}^{-1}$)	Kd_2 (10^{-6})	ΔH_2 (kcal.mol^{-1})	ΔG_2 (kcal.mol^{-1})	ΔS_2 ($\text{Cal.mol}^{-1}\text{K}^{-1}$)
D1D2 + hairpin	0.49 (0.03)	-17.7 ($9.4.10^{-2}$)	-8.6 ($5.5.10^{-2}$)	-30.5 (0.17)	6.2 (0.2)	-8.3 (1.10^{-1})	-7.1 ($2.6.10^{-2}$)	-4.2 (0.3)
D1D2 + duplex	1.0 (0.7)	-20.6 ($1.5.10^{-1}$)	-8.15 (4.10^{-2})	-41.9 (4.10^{-4})	5.7 (0.33)	-2.3 ($1.8.10^{-1}$)	-7.1 ($3.4.10^{-2}$)	16.3 (0.6)

17.2.2.2 Comparison between forward and reverse experiment and tentative of global fitting of both datasets

Fits of the experiments with pre-miR-155 injected into TRBP-D1D2 (reverse experiments) lead to dissociation constants of 0.5 μM and 4 μM for the first two dissociation constants, consistent with the results of the experiment with TRBP-D1D2 injected on pre-miR-155. However the two sum of the two successive reaction enthalpies are about -30 kcal.mol^{-1} (**fig. 17.4**), different from the one obtained from the fit of experiments realized with TRBP-D1D2 in the syringe of about -25 kcal.mol^{-1}. This difference did not enable to have a convergent fit using both forward and reverse experiments. As both forward and reverse experiments are consistent with 2 protein binding one RNA, the results of the forward experiments performed with as many injection as possible in order to progressively form the complexes (adding proteins on the RNA) were preferred for the estimation of the binding parameters. Additionally, the dilution of a highly concentrated RNA solution in reverse experiments could explain the difference observed effect as the RNA has a potential small tendency to self interact. Anyhow, the main goal of the ITC experiment being the estimation of the dissociation constant and possible molar ratios and not the enthalpies of reaction, this choice does not lead to significant differences. Since the larger detected complex formed has a protein/RNA molar ratio of 2, a more complex binding pathway with possible sides effects linked to the kinetic is possible when the experiment starts with this complex and goes towards a complex of protein/RNA ratio of 1. Moreover, in the reverse experiments, pre-miR-155 is used in the syringe at high concentration and the potential small effect linked to the dimerization of the RNA could lead to an additional dilution reaction contributing to the heat transfer providing the equilibrium is fast, and leading to errors in the measurement. For this reasons the reverse experiment is considered less relevant than the forward experiments.

17.2.2.3 Comments of potential microscopic mechanisms

The macroscopic model parameters suggest a higher affinity for the first equilibrium than for the second one. Importantly, this does not directly correspond to the affinity of given sites on the macromolecules. Fitting of a microscopic (i.e. site specific) model was not attempted for the reasons developed in the methodological part and as it would not enlighten the situation at this stage of the study. Weak explanations can be proposed to account for the differences of macroscopic affinities. For two identical independent binding sites on the RNA, two different macroscopic constants are expected due to the difference of the number of free sites at the two successive binding steps, following $\frac{Kd_2}{Kd_1} = 4$. Here the $\frac{Kd_2}{Kd_1} \approx 6$ for pre-miR-155/pre-miR-155* and 12 for pre-miR-155, which is slightly higher than what is expected for the reference model of independent sites and would compatible with a slight negative cooperativity on the two proteins binding the RNA. However as underlined in the ITC chapter, many other more complex microscopic mechanism could lead to the same pair of macroscopic dissociation constants, like an effect on the RNA in the first binding event triggering an RNA structural/dynamical change enabling another TRBP-D12 to bind to a newly favorable site, that couldn't binds TRBP-D12 without the first binding event (example of positive cooperativity).

17.3 ITC analysis of TRBP-D1 & -D2 interaction with pre-miR-155

In order to get insight on the behavior of each dsRBDs in TRBP-D1D2 in the 2:1 complex previously detected, additional ITC experiments were performed with TRBP-D1 and TRBP-D2.

17.3.1 ITC datasets

An example of a typical forward and reverse ITC data obtained with TRBP-D2 without baseline correction are given in **fig. 17.3**. The same kind of data was recorded for TRBP-D1. Contrary to the experiments with TRBP-D1D2 the forward experiments generated signal with a very non flat baseline, since the power signal never comes back to the baseline during the experiments. The importance of this process led to the conclusion that no relevant fitting could be extracted from these experiments. Control with injection of TRBP-D2 in the buffer with the same conditions than in the forward experiment does not show peaks that could indicate homo-interaction (**fig. 17.3**). This is consistent with the analysis of the apo constructs of TRBP. The strong decrease of the baseline when TRBP-D2 is injected on the RNA could be explained by a very slow kinetics or more likely by the formation of some aggregates in solution. Multiple experiments performed with both TRBP-D1 and TRBP-D2 shown that this phenomenon was reproducible and not dependent on the concentration in the range tested. As a result, only the reverse experiment could and were analyzed. They were reproducibly characterized by a common baseline, flat or slightly drifted as in **fig. 17.3** A.. A direct consequence of this is that the comparison of these binding events with the single domains and the double domain construct required to use the reverse datasets of TRBP-D1D2.

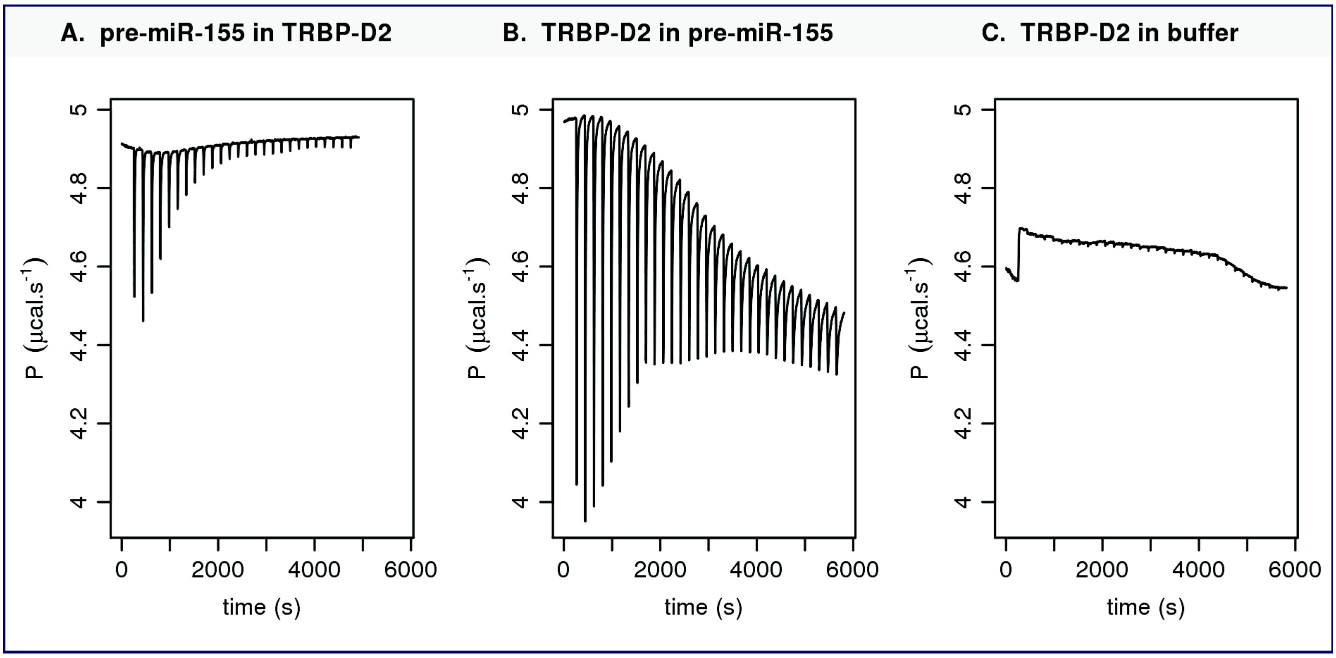


Figure 17.3: Examples of ITC datasets of TRBP-D2 recorded on ITC-200 (MicroCal) at 25°C, without baseline correction. **A.** Experiment with a solution of pre-miR-155 at 150 μM injected into a solution of TRBP-D2 at 36 μM with 26 injections of 1.5 μl lasting 3 seconds and spaced by 3 minutes intervals. **B.** Experiment with a solution of TRBP-D2 at 1.2 mM injected into a solution of pre-miR-155 at 35 μM with 26 injections of 1.25 μl lasting 2.5 seconds and spaced by 3 minutes intervals. **C.** Control experiment with a solution of TRBP-D2 at 400 μ injected into the phosphate buffer solution of pre-miR-155 at 35 μM with 26 injections of 1.25 μl lasting 2.5 seconds and spaced by 3 minutes intervals.

17.3.2 Application of the fitting procedure on TRBP-D1 and TRBP-D2

Fitting of the ITC data of the isolated first two dsRBD of TRBP The application of macroscopic models of the reverse experiments with TRBP-D1 and TRBP-D2 did not lead to good fits with protein: RNA molar ratio of 1 and 2 and higher models had too much parameters for the few features of the datasets. Thus an accurate determination of the molar ratio and binding affinities could not be estimated with macroscopic models. Exceptionally, TRBP-D1D2, TRBP-D1 and TRBP-D2 ITC data interaction with pre-miR-155 were fitted with a model of single independent sites with a floating molar ratio parameter. As already mentioned, the strong assumptions of this model are absolutely not guaranteed in the present case. The use of this simplified model enabled however a global comparison of the binding properties of the three proteins constructs under the same the hypothesis of independent binding domains. The fitting results are given in **fig. 17.4**. Application of this model led a molar ratio of 2 for TRBP-D1D2 as expected, and a molar ration of 4 for TRBP-D1 and TRBP-D2 meaning that 4 dsRBDs are able to bind to a single molecule of pre-miR-155, a result that seems reliable even with the assumption used. The binding constant however provides information of the global affinity of the TRBP constructs and suggests a higher affinity for the second dsRBD of TRBP compared to the first one. The result for TRBP-D1D2 can be compared to the ones obtained without the site identicity and independency. Under these hypothesis of identical and independent sites, the microscopic constant K_{micro} can be used to express the two related successive macroscopic constants $K_{\mathrm{d},1}$ and $K_{\mathrm{d},2}$ with $K_{\mathrm{d},1} = 0.5K_{\mathrm{micro}}$ and $K_{\mathrm{d},1} = 2K_{\mathrm{micro}}$, leading to $K_{\mathrm{d},1} = 0.510^{-6}$ and $K_{\mathrm{d},1} = 6.810^{-6}$. Despite the absence of guaranties of the aforementioned hypotheses, these values are very similar to the ones obtained without these strong hypotheses (0.49 10^{-6} and 6.2 10^{-6} respectively), suggesting that is approach is not completely unrealistic for the system under study. The results of the fitted parameters are provided in **tab. 17.4**. Under the hypothesis of identical independent sites, these microscopic constants correspond to the following successive macroscopic constants: 3.8 10^{-6}, 10.3 10^{-6}, 23.1 10^{-6} and 61.6 10^{-6} for TRBP-D1 and 1.0 10^{-6}, 2.7 10^{-6}, 6.1 10^{-6} and 16.4 10^{-6} for TRBP-D2.

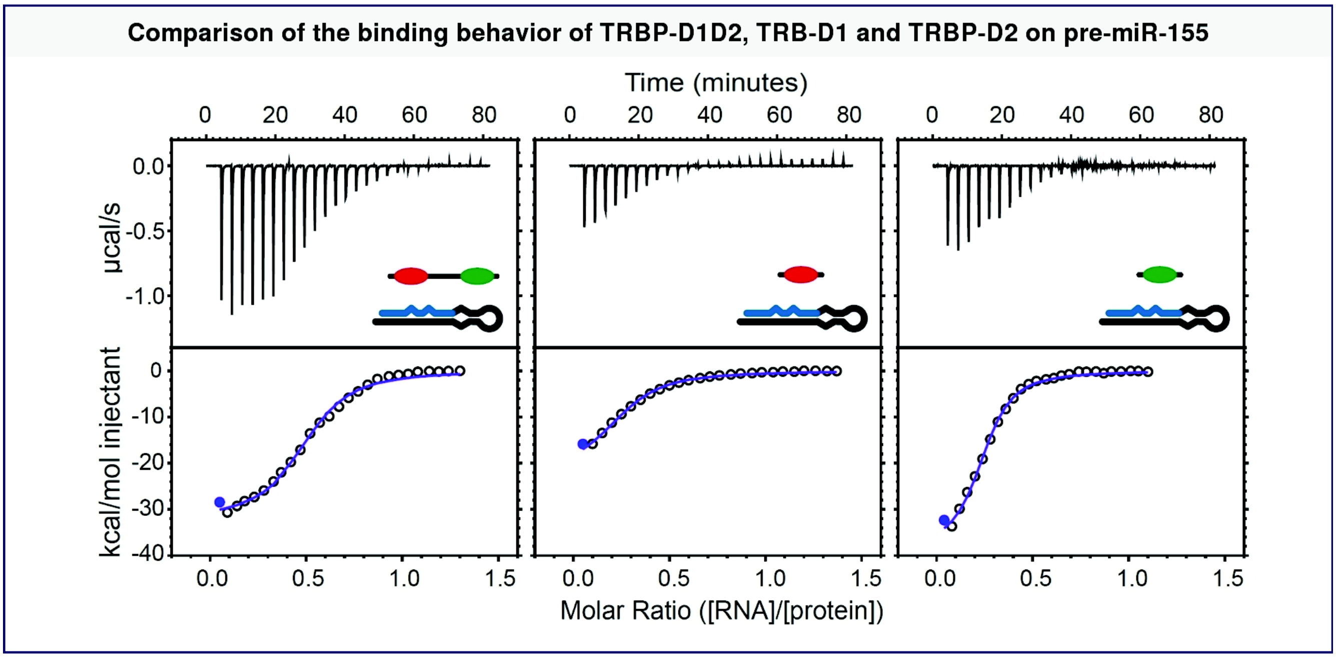

Comparison of the binding behavior of TRBP-D1D2, TRB-D1 and TRBP-D2 on pre-miR-155

Figure 17.4: Comparison of the binding behavior of TRBP-D1D2, TRBP-D1 and TRBP-D2 on pre-miR-155 using ITC experiments with pre-miR-155 injected into TRBP protein constructs recorded on ITC-200 (MicroCal) at 25°C. Data fitting for this comparison is performed under the hypothesis of a single type of independent sites on pre-miR-155. From left to right, pre-miR-155 (160 μM, 300 μM, 300 μM) injected into a solution of TRBP-D1D2 (50 μM), TRBP-D1 (45 μM) and TRBP-D2 (20 μM). Experiments were performed with 26 injections of 1.5 μl lasting 3 seconds and spaced by 3 minutes intervals.

Table 17.4: Result of the estimation of thermodynamic parameters for the interaction between TRBP-D1D2 or TRBP-D1 or TRBP-D2 with pre-miR-155. This estimation is performed only for comparison purpose under the strong hypothesis of independent identical binding sites on pre-miR-155. The standard deviations of each parameter are indicated into parenthesis, using fitting from two experiments per system.

System	Kd (10^{-6})	Molar ratio	ΔH (kcal.mol^{-1})	ΔS (Cal.mol^{-1}K^{-1})
pre-miR-155 + TRBP-D1	15.4 (0.8)	3.9 (0.2)	6.0 (0.2)	1.9 (0.2)
pre-miR-155 + TRBP-D2	3.7 (0.5)	4.1 (0.1)	9.4 (0.2)	6.8 (0.2)
pre-miR-155 + TRBP-D1D2	3.4 (0.8)	2.0 (0.1)	16.1 (0.3)	28.8 (0.3)

17.4 Conclusion of the ITC analysis

Under the probable hypothesis of identity of the molar ratio with the stoichiometry and on the absence of undetected interactions, the ITC analysis shown that two RNA binding region of TRBP are able to bind with a similar affinity with the miR-155 precursor product (pre-miR-155) and substrate (miR-155/miR-155*) of Dicer. The affinity of the binding events is around 10^{-6}. The aforementioned hypothesis will be test in the next chapter by svAUC. Both isolated dsRBDs are able to interact with pre-miR-155, with and higher affinity for the second dsRBD which is close to the affinity of TRBP-D1D2 under the hypothesis of a single type of independent interacting sites on the RNA. The data do not enable a reliable study of microscopic mechanism of this interaction (*e.g.* cooperativity).

Various mechanisms could be compatible with the 2 TRBP-D1D2: 1 RNA molar ratios. For the simplest case with a stoichiometry of 2, it can be conceived :

- only one of the two TRBP domains binds in the two successive events, either the same of two different domains, without or without a contribution to the interdomain region. The fact both isolated dsRBDs are able to bind the pre-155 precursors so this seems unlikely at this stage.
- all the domains binds when each molecule bind, without or without a contribution to the interdomain region.
- intermediates between the last two situations.

And the possibilities are higher for more complex stoichiometries. The available data on the complex at this stage of the study does not allow to restrain this set of hypothesis. The steps performed to get additional insight on the interaction are the following ones :

- use of an additional technique to check if the molar ratio is the actual stoichiometry and potential undetected interactions: svAUC
- use of NMR to get site specific information of TRBP-D1D2 binding surfaces, on a system with a protein-RNA ratio well chosen from the stoichiometry and binding constant information.

Chapter 18

Stoichiometry analysis of the RNA binding region of TRBP by svAUC

Requirement of additional experiments to determine the stoichiometry and validate ITC dissociation constants The goal of this chapter is to get an independent validation of the stoichiometry proposed in the ITC study, and as a consequence of the ITC derived binding constants. A molar ratio of 2 TRBP-D1D2 for 1 RNA has been obtained non ambiguously from ITC data. But as mentioned in the ITC part there are multiple phenomena that can give similar results. First there is still the possibility of additional reactions happening without enthalpy change. Then, even in the absence of additional undetected reactions, there are many stoichiometries compatible with protein/RNA molar ratios of value 2 (2 protein / 1 RNA; 4 protein / 2 RNA, etc). The knowledge of the actual stoichiometry is one of the most important parameter to define macroscopically an interacting system. Importantly, no information was reported in the literature about the stoichiometry of TRBP with miRNA precursors. This analysis is performed with TRBP-D1D2 and pre-miR-155.

18.1 svAUC analysis of TRBP with pre-miR-155

18.1.1 svAUC experimental design

Sedimentation velocity experiments were performed as described in the methodological part. The initial estimations of the dissociation constants by ITC have driven the choice of the concentrations for the svAUC experiment. As in the ITC experiment, the protein had to be in excess over the RNA concentration in order to have the complexes of high protein/RNA ratios formed, corresponding to the molar ratio of 2 found in ITC. The svAUC data had also to show the potential existence of a 1 protein: 1 RNA intermediate complex, formed during the formation of the 2 protein: 1 RNA complex, as well as potential other complexes not detected in ITC. These goals have contributed to the experimental design of the svAUC experiment. Additional constraints include the work at low concentration to avoid UV detector saturation, to get close the assumptions used to derive svAUC equations and to avoid non ideal hydrodynamic and thermodynamic effects. It is an advantage to have the RNA at the lowest concentration since it is more prone to non ideal effects and because it absorbs strongly in the UV region.

The study of the isolated components was performed at the same time as the study of the complexes. The results have already been developed in the parts devoted to the apo-TRBP and apo-miRNA-155. The parallel characterization of the free and bound macromolecule was important since the sedimentation parameters of the free species were required to study the complexes. As a result, 3 cells among the 7 ones of the An-50Ti rotor were available for the complexes. Using the pair of dissociation constants estimated in the ITC part, it is possible to estimate the abundance of the potential 1:1 and 2:1 complexes for the selected total concentration of TRBP-D1D2 (P) and pre155 (R) (**tab. 18.1**). Results are given for the conditions selected for the svAUC experiment. These calculations were performed under the hypothesis that the measured binding constant in ITC is equivalent to the real binding constant and that the protein/RNA molar ratios of 1 and 2 corresponds to complexes of {1 protein: 1 RNA} and {2 protein: 1 RNA} respectively. Despite some errors in the ITC measurement have been taken into account, it is likely that this assumption is not perfectly true as already mentioned in the thermodynamical part on the binding constants. As a result, the actual ratios $\frac{[RP_2]}{[RP_1]}$ might not be exactly the ones in solution i.e. in the svAUC cells.

Table 18.1: Prediction of the concentration of the species in solution in the svAUC conditions and assuming a macroscopic equilibrium with two successive addition of TRBP-D1D2 on pre-miR-155. R: pre-miR-155 ; P: TRBP-D1D2.

C(P) μM	**C(R)** μM	**[P]** μM	**[R]** μM	**[RP]** μM	**[RP$_2$]** μM	$\frac{[\mathbf{RP_2}]}{[\mathbf{RP_1}]}$	**sd**$\left(\frac{[\mathbf{RP_2}]}{[\mathbf{RP_1}]}\right)$
10	0.5	$9.2\ 10^{-6}$	$1.1\ 10^{-8}$	$2.0\ 10^{-7}$	$2.9\ 10^{-7}$	1.50	0.11
10	1	$8.4\ 10^{-6}$	$2.4\ 10^{-8}$	$4.1\ 10^{-7}$	$5.6\ 10^{-7}$	1.37	0.10
10	2	$7.0\ 10^{-6}$	$6.4\ 10^{-8}$	$9.0\ 10^{-7}$	$1.0\ 10^{-6}$	1.15	0.08

The Monte Carlo simulation done to estimate the parameter errors in the ITC fitting can be also used to estimate the distribution of $\frac{[RP_2]}{[RP_1]}$. This shows that 95% of the values are within the mean +/- two standard deviations. The values of the expected $\frac{[RP_2]}{[RP_1]}$ ratios are given in **tab. 18.1**. As a result, the conditions used appeared to a be a compromise between the different constraints previously mentioned. Another possible design of this experiment would have been to add variable amounts of proteins to a constant low RNA concentration. A more targeted demonstration of the existence of the potential {2 protein:1 RNA} complex would have been possible using a much higher amount of protein (100:1) in order to have almost no contribution of a potential 1:1 complex.

18.1.2 Data analysis with a $c(s)$ model

$c(s)$ fitting As explained in Chapter 11, the partial specific volumes for possible protein/RNA complexes were estimated by summing the individual values of each component weighted by their fractional mass in the complex. For {1 TRBP-D1D2: 1 pre-miR-155 }, {2 TRBP-D1D2: 1 pre-miR-155 } and {1 TRBP-D1D2: 2 pre-miR-155 } complexes this led to partial specific volumes of about 0.41ml/g, 0.64 ml/g and 0.43 ml/g. $c(s)$ profiles reported here were fitted with a partial specific volume value of 0.64 ml/g. Importantly, the sedimentation coefficient, s, of each complex estimated from the distribution of sedimentation coefficients, $c(s)$, was not affected by the choice of the 3 possible $\overline{v}$ values. The overlaid $c(s)$ profiles for the different cells are given in **fig. 18.1**.

Sedimentation coefficients analysis SvAUC data acquired at three concentrations with an excess of protein in **fig. 18.1** show the disappearance of the RNA contribution at 4.4 S and the concomitant emergence of a new faster contribution, which sediments at about 5.9 S. An additional minor contribution (6 % of the main peak) is also detected at 7.5 S only in the sample with 2 μM RNA, at almost the same position than the probable RNA dimer detected in the experiment with pre-miR-155 alone at 2 μM. The nature of this peak will be discussed later, in the buoyant molar mass analysis.

Three concentrations of RNA were used to try to estimate the kinetic regime of the exchange on the svAUC time scale. The different kinetics regimes were described in **eq. (11.9)**. There is not a clear tendency in the relative peak position of the main peak as a function of concentration. However, only three concentrations have been used so it is not possible to discriminate between positions changes caused by errors in the data or real effects of the concentration. The lowest concentrated sample with 0.5 μM of RNA is the most likely to have a significant error contribution in the $c(s)$ analysis. Despite this we see that in the range of concentration sampled, the s-value for the main peak is almost concentration invariant. Because of the detection of at least two complex species in ITC, the peak at 5.9 S should corresponds to a reaction boundary that contains a mixture of oligomeric species (242; 243; 303). This is consistent with the behavior of complex exchange on fast or intermediate svAUC regime. For a reaction boundary, it is predicted that the sedimentation coefficient and apparent buoyant molar masses determined from the diffusion broadening of the fast boundary, would be less than the value expected for the largest possible complex formed (242; 243; 303).

In order to understand the stoichiometries that can lead to the observed sedimentation coefficient values, the experimental values resulting from the $c(s)$ analysis were compared to theoretical ones obtained for different protein/RNA ratios. This calculation requires the prior knowledge of the unknown frictional ratio of the complexes. Since both the free protein and the free RNA has a frictional ratio of about 1.4, this value was used in this modeling as well as a value of 1.25 corresponding to a spherical complex. Values of frictional ratio higher that 1.4 would imply the formation of a complex more elongated than the individual partners. It is perfectly conceivable and will be discussed later. The results of these calculations are summarized in **tab. 18.2** and the corresponding range of sedimentation coefficient is represented in **fig. 18.1**.

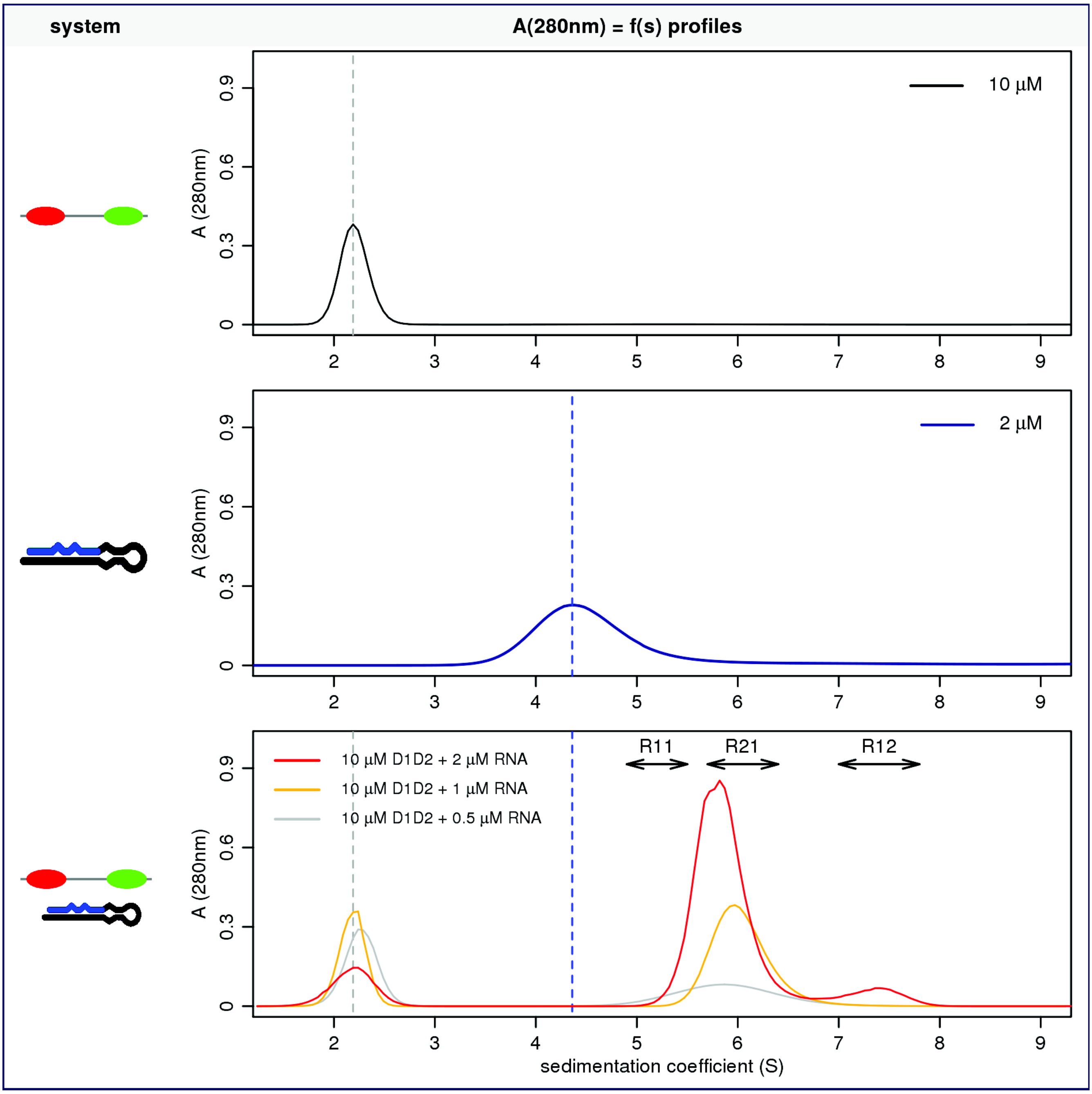


Figure 18.1: Comparison of the diffusion deconvoluted sedimentation coefficient distribution of TRBP-D1D2 (10 μM), pre-miR-155 (1 μM) and 3 mixtures of these two macromolecules at 25°C. The region R11, R21 and R12 correspond to the expected position of 1:1, 2:1 and 1:2 TRBP-D1D2:pre-miR-155 complexes under the hypothesis of slow exchanging complex on the svAUC time scale. The range of sedimentation coefficient reflects the uncertainty of the frictional ratio of these complexes. The region starts from a frictional ratio of 1.4 (similar to the asymmetric isolated TRBP-D1D2 and pre-miR-155) to 1.25, corresponding to a spherical complex.

The s-value determined for the fast boundary in **fig. 18.1** is larger than that expected for a 1:1 TRBP-D1D2/pre-miR-155 complex, for the frictional ratio 1.4 and 1.25 and very close to the one expected for a reaction boundary with a high proportion of 2:1 complex. Considering the calculated s-values for possible globular protein/RNA complexes with stoichiometries of 1:1, 2:1 and 1:2 in **tab. 18.2**, only the complex composed of two proteins and a single RNA is close to the experimentally determined value. The other possible complex compositions differ by more than 1 S. As a consequence these results are not compatible with the exclusive presence of protein/RNA complexes whose stoichiometry is a multiple higher than 1 of the protein over RNA molar ratios of 1 and 2 estimated with ITC. The minor peak at 7.5 S has a sedimentation coefficient similar to the one of a complex associating two RNA and a protein with a globular shape.

Table 18.2: Prediction of the concentration of the species in solution in the svAUC conditions and assuming a macroscopic equilibrium with two successive addition of TRBP-D1D2 on pre-miR-155. R: pre-miR-155; P: TRBP-D1D2.

Experimental values						Calculated values							
System	s_{exp} (S)	M_b (kDa)	$\frac{A_{260}}{A_{280}}$	$\frac{\Delta J}{A_{280}}$	$\frac{\Delta J}{A_{260}}$	P:R	M_{calc} (kDa)	$\bar{v}$ (ml.g^{-1})	M_{calc} (kDa)	$M_{b,calc}$ (kDa)	$\frac{A_{260}}{A_{280}}$	$\frac{\Delta J}{A_{280}}$	$\frac{\Delta J}{A_{260}}$
D1D2	2.1(0.1)	5.7	1.73	5.7	9.9	1:0	22.3	0.74	2.1	5.7			
pre155	4.4(0.2)	9.1	0.46	0.3	0.14	0:1	19.6	0.508	4.3	9.2			
						1:1	41.9	0.64	4.9-5.5	14.9	0.47	0.53	0.25
Complex	5.9(0.1)	18.3	0.49	0.8	0.41	2:1	61.5	0.67	5.7-6.4	20.6	0.49	0.75	0.37
						1:2	64.2	0.60	7.0-7.8	24.0	0.47	0.42	0.20

18.1.3 Data analysis with Lamm equation modeling of non interacting species

Buoyant mass analysis Information can be obtained on reaction boundaries considering them as a non interacting species (303). The estimated protein buoyant molar mass is 4.4 ± 0.4 kDa instead of 5.7 kDa for the monomer. The apparent buoyant mass estimation for the fast boundary is 18.3 ± 1.0 kDa, which is between the values calculated for a 1:1 or 2:1 TRBP-D1D2/pre-miR-155 complex and close to the 2:1 complex. Thus the non interacting species model yields under-evaluated buoyant masses. This is what is expected for a system with both 1:1 and a larger complex in equilibrium. Thus this analysis shows the existence of the 1:1 complex but does not add key information on the larger one(s). The buoyant molar mass calculated for the minor species ($s = 7.5$ S) was similar to that of the main complex, thus irrelevant with the data at hand. This peak could correspond either to the potential RNA dimer detected on the free RNA, either to an artifact linked to deconvolution used in the $c(s)$ analysis which requires high quality. The buoyant mass analysis gives with no surprises less information that the comparison of the sedimentation coefficients with theoretical ones.

Multiple wavelength analysis To further characterize the complexes contributing to the boundaries, a composition analysis of the complex using the data obtained from the 3 different optics used was performed. This analysis based only on signal intensity and not on the sedimentation parameters is independent of the sedimentation coefficient and buoyant mass analysis presented before. The extinction coefficients at 280 nm for TRBP-D1D2 (10,084 cm^{-1} mol^{-1}) and at 260 nm for pre-miR-155 (490,000 cm^{-1} mol^{-1}) were fixed in order to derive the other extinction coefficients (at 280 nm and 260 nm) and increments of refractive index ($\frac{dn}{dc}$). The sedimentation profiles obtained for the three optics and the $c(s)$ distributions were simultaneously analyzed for each of the free compounds using a hybrid model with three non-interacting species. The resulting ratios of absorbance and interference fringes are given in **tab. 18.2**. These ratios were used to calculate values for different protein/RNA stoichiometries. The experimental ratios measured in the $c(s)$ analysis for the fast boundary reflects an enrichment of protein in the complex characterized by about 2 times more protein signal than RNA signal. This ratio is higher than the one expected from the experimental design of the svAUC experiment under the assumption than the ITC dissociation constants match the true constants and with the simplest stoichiometries compatible with the molar ratio. The observed difference might be linked to the activity and additional small species contributions to the actual binding constant presented, caused by ITC biases (co-reactions for instance) or because of the presence in solution of a species in equilibrium with the other ones characterized by a protein/RNA molar ratio higher than 2. These hypotheses are all compatible. This situation will be clarified in the next paragraphs.

Conclusion: svAUC contribution to the stoichiometry information The diffusion deconvoluted sedimentation coefficient analysis ($c(s)$) and modeling of the sedimentation data by Lamm equations assuming the species in solution sediments like non interacting ones (buoyant mass analysis and multiple wavelength analysis) gives sedimentation parameters that are intermediate between a 1:1 complex and another complex enriched about two times in proteins, the smaller conceivable one being a {2 protein: 1 RNA complex}. Because of the properties of the reaction boundaries, this show the (unsurprising) existence in solution of a 1:1 stoichiometric protein/RNA complex (intermediate), expected from the first binding event detected in ITC. The svAUC data analyzed at this step do not enable to go further. Various increasingly complicated hypotheses are compatible to them regarding the complexes of molar ratio higher than 1 :

- Among the complexes with TRBP-D1D2/pre-miR-155 molar ratio over 1, only the complex with 2 molecules of TRBP-D1D2 and one pre-miR-155 exists. This would be compatible with all the svAUC data and ITC data. This complex is formed by the successive interaction of 2 TRBP-D1D2 on a pre-miR-155.
- Among the complexes with protein/RNA molar ratio over 1, only complexes of molar ratio 2

exists (2 TRBP-D1D2 with 1 pre-miR-155 or 4 TRBP-D1D2 with 2 pre-miR-155, etc). This would be compatible with all the svAUC data and ITC data but would require to adapt the estimation of the ITC dissociation constants.

- Among the complexes with protein/RNA molar ratio over 1, complexes of molar ratio 2 and more exists. Molar ratios higher than 2 would require reaction enthalpies of almost 0 in the ITC condition used (i.e. not detected in ITC).

The last two hypothesis seem very unlikely.

18.2 Simulation of svAUC data using the ITC data as knowledge

The experimentally estimated sedimentation coefficient and buoyant masses of the reaction boundary are larger than the ones expected for the existence of only 1:1 complexes, indicating the existence of a larger complex. This deduction was obtained from the successive analysis of the svAUC data. To complete the analysis and prove the overall consistency of the reasoning, prediction of the svAUC data can be attempted using the available information. Under the hypothesis that the ITC dissociation constants are good estimates of the constants of the equilibrium involved in solution, implying that there is both 1:1 and 2:1 stoichoimetric TRBP-D1D2:pre-miR-155 complexes, do the information obtained from the svAUC analysis combined with ITC dissociation constants enable to reproduce the svAUC data ? As svAUC raw sedimentation data are not easily directly exploitable, the diffusion deconvoluted analysis was use to analyze the simulated datasets. The goal was not to simulate exactly the experimental $c(s)$ curves since the frictional ratios of the complexes are not known, but to test the compatibility of the experimental data in reasonable range of the frictional ratios. Without enabling to clearly select one of the three aforementioned hypotheses, good prediction of the sedimentation data would give an additional (weak) argument to the existence of these two simple complexes compared to the more other alternatives. On the other hand these simulations enable to check whether the dissociation constants reported in ITC are consistent with the svAUC data, especially regarding the number and shapes of peaks observed in svAUC. Two strategies were used :

- The sedimentation coefficients and buoyant mass values observed experimentally can be predicted using numerical simulations of Lamm equations with non null chemical reaction flux between the species followed by a diffusion deconvoluted $c(s)$ analysis. This required previous knowledge on dissociation constants and hypothesis on kinetics constants.
- Under the assumption of a fast exchange on the svAUC time scale, Schuck and colleagues have investigated the analysis of bimodal boundaries (like the one observed here) with Lamm equation modelization of non-interacting species analysis. They term the fast boundary an "effective particle" (243; 242; 303). This "effective particle" model proved to be also useful for other exchange regimes.

The results of both approaches are reported in the next two sections.

18.2.1 Simulation protocol for Lamm equation modelization of TRBP-D1D2 / pre-miR-155 equilibrium

18.2.1.1 Description of svAUC simulation with numerical Lamm equation

The simulations were performed for a model {R + 2P = RP + P = RP$_2$}, with R being the RNA pre-miR-155 and P the Protein TRBP-D1D2. The conditions correspond to pure water at 20°C (s_{20w}; $\rho = 0.99828$ g.ml^{-1} and $\eta = 1.002$ cp), required by the Sedphat program (10). The program Sedfit was initially used to generate 60 svAUC profiles for the sedimentation at 50,000 rpm of a molecule at a concentration $c = 1.0$ (arbitrary unit), from radial distances of 6.0 to 7.2 cm, in 600 second increments, with an artificially generated Gaussian noise of standard deviation 0.01 arbitrary unit. For these simulations we used $s = 4$ S and $D = 5.10^{-7}$ cm^2 s^{-1}. The noise, c, s and D values are arbitrarily defined and did not impact the analysis below. The experimental conditions defined by the set of profiles generated in this way were then used in the program Sedphat to generate sedimentation profiles corresponding to the association equilibrium under study. The s_{20w} profiles were calculated with a frictional ratio of 1.4. The partial specific volume $\overline{v}$ for all species was 0.75 ml g^{-1}, so that apparent molar masses used were $M_{app} = \frac{M_b}{1-\rho\overline{v}}$ with M_b the actual buoyant molar masses. The simulation was performed using extinction coefficients at 280 nm. A summary of the parameters used in the simulations is given in **tab. 18.3**.

Table 18.3: Summary of shared parameters used in the simulations to model svAUC profiles using Sedphat (10)

Molecule	s_{20w}	M_b (kDa)	M_{app} (kDa)	$\epsilon_{280\text{nm}}$ (cm^{-1} M^{-1})
R (pre-miR-155)	3.84	9.6	38.177	225975
P (TRBP-D1D2)	1.92	5.7	22.745	10084
RP	4.45	15.3		
RP$_2$	5.18	21.0		

Three kinetic regimes were used for each of the 3 concentration conditions used in the experimental part. Simulations groups 1-4, 5-8 and 9-12 show the effect of varying the association kinetics from fast to slow, i.e. from conditions under which a single effective particle would be seen to conditions where two sedimenting boundaries for species RP and RP$_2$ would be seen. Exaggeratedly slow off rates have been used in the latter case. In these groups, the dissociation constants used in the simulation corresponds to the dissociation constants that reflect those determined by ITC apart from the simulations 3, 7 and 11 that were performed to see the effect of equal macroscopic constants on the svAUC profiles. Kinetics faster than the fastest regime and lower than the lowest one are expected to give no distinguishable changes in the svAUC profiles. In each simulation group 1-4, 5-8 and 9-12, the concentrations of protein and RNA are varied to match the range used in our experimental study. Finally, simulations 13-15 reflect conditions in which the equilibrium is skewed towards the formation of a 1:1 complex, RP (simulations 13 and 14) or the 2:1 complex, RP_2 (simulation 15). These simulations establish the theoretical s-values of RP and RP_2 complexes.

Table 18.4: Concentrations of pre-miR-155 (R) and TRBP-D1D2 (P), equilibrium constants and kinetic dissociation constants used in the simulations.

Simul. Name	[R] (μM)	[P] (μM)	$log(Ka_1)$	$log(k_{-1})$	$log(\frac{Ka_2}{Ka_1})$	$log(\frac{k_{2-}}{k_{-1}})$	Kd_1	Kd_2	k_{-1} (s^{-1})	k_{-2} (s^{-1})
1	0.5	10	6.3	-1	-1	$3.01.10^{-1}$	0.5	5.0	0.1	0.2
2	1	10	6.3	-1	-1	$3.01.10^{-1}$	0.5	5.0	0.1	0.2
3	1	10	5.3	-1	0	$3.01.10^{-1}$	5.0	5.0	0.1	0.2
4	2	10	6.3	-1	-1	$3.01.10^{-1}$	0.5	5.0	0.1	0.2
5	0.5	10	6.3	-3	-1	$3.01.10^{-1}$	0.5	5.0	1.10^{-3}	2.10^{-3}
6	1	10	6.3	-3	-1	$3.01.10^{-1}$	0.5	5.0	1.10^{-3}	2.10^{-3}
7	1	10	5.3	-3	0	$3.01.10^{-1}$	5.0	5.0	1.10^{-3}	2.10^{-3}
8	2	10	6.3	-3	-1	$3.01.10^{-1}$	0.5	5.0	1.10^{-3}	2.10^{-3}
9	0.5	10	6.3	-6	-1	$3.01.10^{-1}$	0.5	5.0	1.10^{-6}	2.10^{-6}
10	1	10	6.3	-6	-1	$3.01.10^{-1}$	0.5	5.0	1.10^{-6}	2.10^{-6}
11	1	10	5.3	-6	0	$3.01.10^{-1}$	5.0	5.0	1.10^{-6}	2.10^{-6}
12	2	10	6.3	-6	-1	$3.01.10^{-1}$	0.5	5.0	1.10^{-6}	2.10^{-6}
13	2	10	6.3	-6	-3	$3.01.10^{-1}$	0.5	5.102	1.10^{-6}	2.10^{-6}
14	2	10	6.3	-6	-6	$3.01.10^{-1}$	0.5	5.105	1.10^{-6}	2.10^{-6}
15	2	10	9	-6	-1	$3.01.10^{-1}$	1.10^{-3}	0.0	1.10^{-6}	2.10^{-6}

The generated sedimentation profiles were saved with a Gaussian distributed noise of standard deviation 5 10^{-3} arbitrary unit (c = 1 arbitrary unit). The simulated data were then analyzed with a diffusion deconvoluted $c(s)$ analysis and fit of numerical Lamm equations of non-interacting species. In the latter the fast boundary is treated as an effective particle (303; 242), even if it corresponds to the different interacting species and complexes.

18.2.1.2 Results of the svAUC simulations

The results of the simulations are plotted on the two following figures.

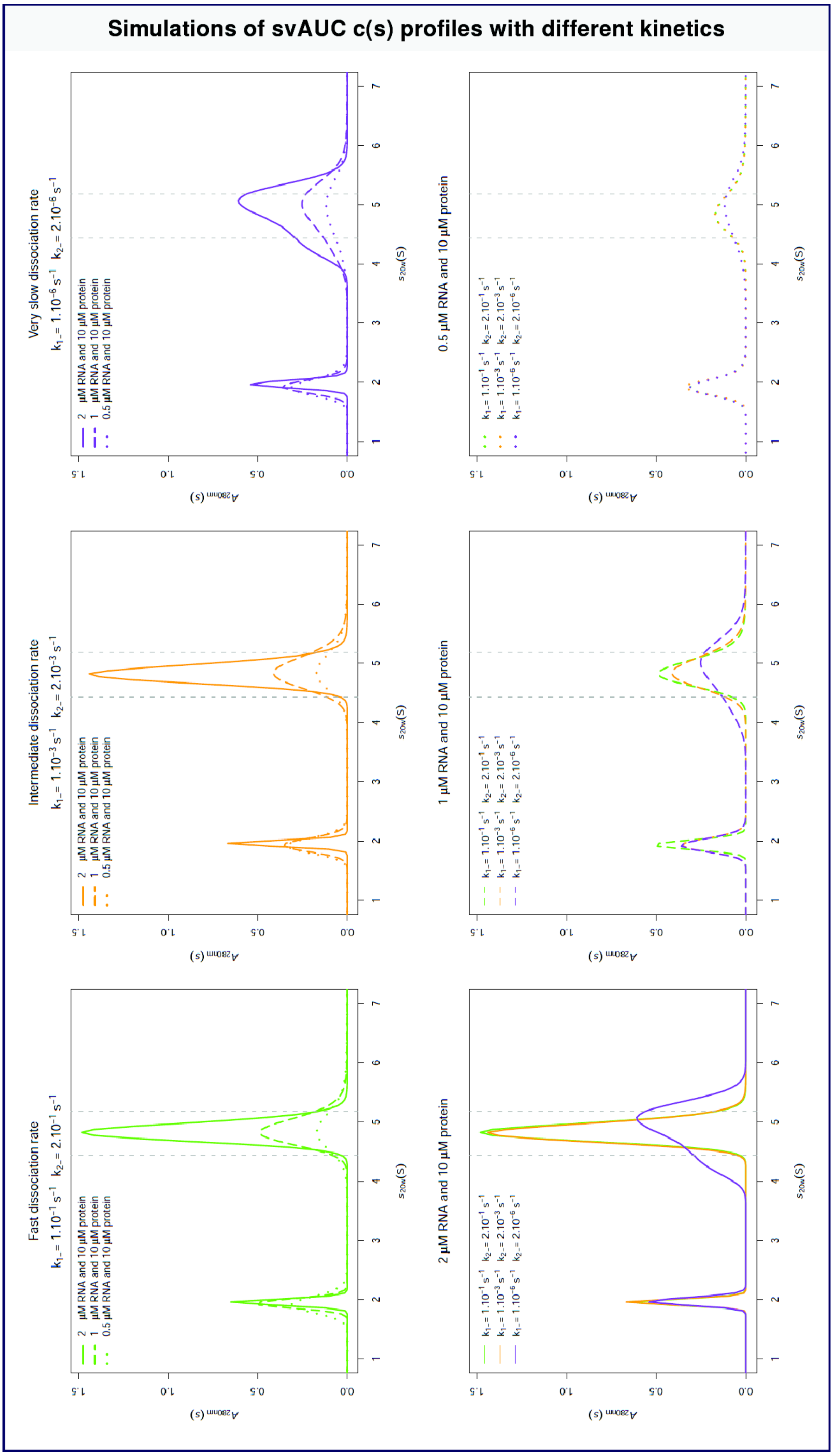


Figure 18.2: Effect of the association/dissociation kinetics on svAUC profiles generated with the ITC dissociation constants for different concentrations corresponding to the ones used in the experiments. The vertical dashed lines indicate the theoretical position of the 1:1 and 2:1 protein RNA complexes if there would be only one type of complex in solution.

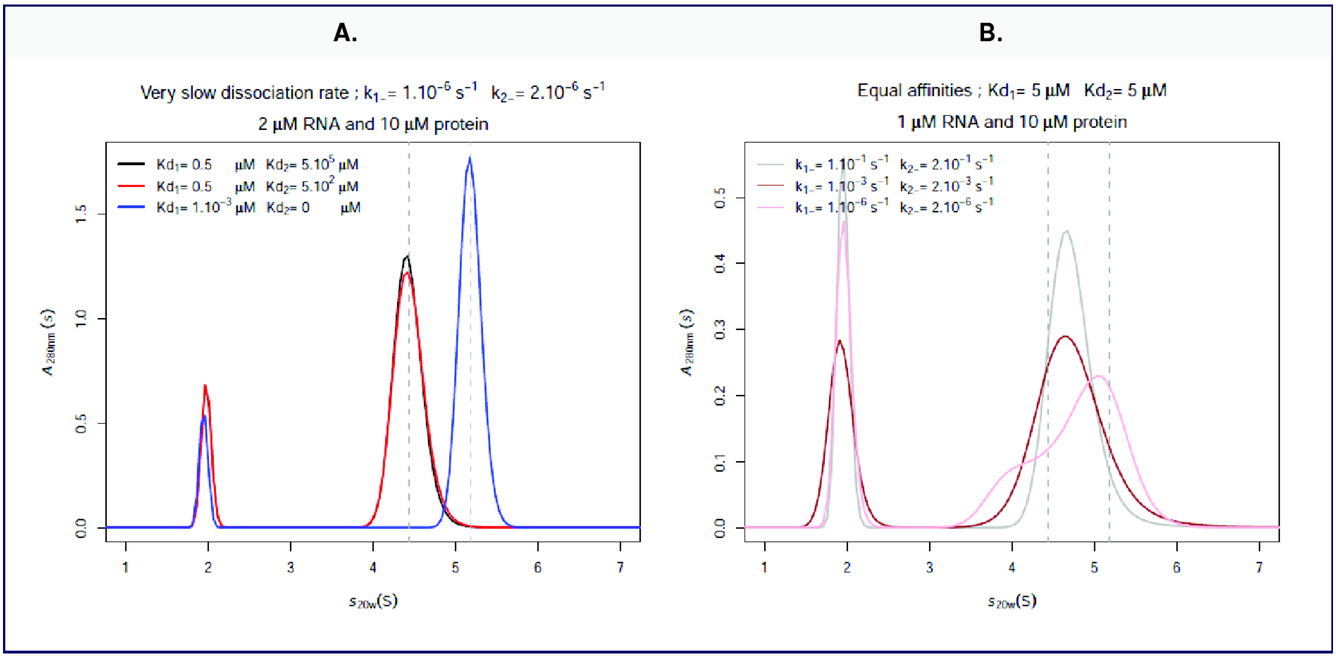


Figure 18.3: **A.** Effects of varying the dissociation constants to favor 1:1 or 2:1 protein/RNA complexes. **B.** Kinetic effect on the svAUC profile with equal macroscopic dissociation constants. The vertical dashed lines indicate the theoretical position of the 1:1 and 2:1 protein RNA complexes if there would be only one type of complex in solution.

Modification of the association kinetics from fast to slow causes an asymmetric enlargement of the fast peak (at about 4.9 S) in the $c(s)$ distribution **fig. 18.2**, first row). The shape of the $c(s)$ distribution is only slightly modified as a function of the RNA concentration. Apart from the case of a very slow kinetic, the peaks shapes are very similar **fig. 18.2**. A numerical summary of the three sets of $c(s)$ distributions and non-interacting species analysis are provided in the two tables below. s_{20w} values for the two peaks seen in the simulated $c(s)$ distributions are given in **tab. 18.5** and the result of Lamm fitting for a non interacting system is given in **tab. 18.5**.

For all the simulated data sets, in the $c(s)$ and the non-interacting species analysis, the slow component has a s-value close to that expected for the pure monomeric protein (s_{20w} value: 1.92 S). The buoyant mass M_b in the model of non-interacting species is also close to the expected M_b for this species (5.7 kDa). The s-value of the fast peak in simulations which uses equilibrium constants that reflect those we determined by ITC, is $s_{20w} = 4.9S$. This value is intermediate between that of RP (4.45 S) and RP_2 (5.18S). The s-value of the fast peak, for simulations 13-14 and 15 corresponds as expected to the inputs RP and RP_2, respectively. For simulations which corresponded to the equilibrium constants determined for our system by ITC and with the fastest kinetics rates (simulations 1,2 and 4), the M_b-value of the fast peak, is about 18 kDa which matches the experimental value. This value is intermediate between the inputs of the pure RP (15 kDa) and RP_2 (21 kDa). The calculated M_b is decreased for simulations corresponding to slower kinetics. In simulations designed to skew the equilibrium towards RP

Table 18.5: Summary of $c(s)$ analysis performed on the simulated datasets. s_{20w} values are those expected for a frictional ratio of 1.4. M_b is the buoyant molar masses.

	Slow component			**Fast component**		
Simulation	s_{20w}	signal	%	s_{20w}	signal	%
1	1.92	0.11	43	4.90	0.14	56
2	1.94	0.10	26	4.88	0.29	73
3	1.94	0.10	27	4.73	0.28	73
4	1.94	0.08	13	4.83	0.57	87
5	1.93	0.11	43	4.90	0.14	56
6	1.94	0.10	26	4.88	0.29	73
7	1.93	0.10	27	4.70	0.28	73
8	1.96	0.08	13	4.83	0.57	87
9	1.93	0.11	44	4.93	0.14	56
10	1.94	0.10	26	4.91	0.29	73
11	1.96	0.11	28	4.78	0.28	72
12	1.97	0.09	13	4.87	0.57	87
13	1.98	0.10	16	4.44	0.56	84
14	1.98	0.10	16	4.43	0.56	84
15	1.94	0.07	11	5.18	0.59	89

or RP_2, respectively, the M_b is close but slightly smaller than the input value.

18.2.1.3 Conclusions

The simulations reported here resemble strongly the peak shapes observed in the experimental svAUC data, especially for fast and intermediate kinetic regimes. Even under conditions of extremely slow off rates, they is only a small shoulder on the fast reaction boundary in the $c(s)$ profile, due to the coexistence of 1:1 and 2:1 species under these conditions. In all other cases, the $c(s)$ profile is characterized by a single peak corresponding to an effective particle that sediments faster than an ideal 1:1 species but slower than a 2:1 species. The exact position of the fast reaction boundary is, obviously, dependent on the kinetic and thermodynamic parameters input into the simulation. It is important to underline here that the frictional ratios of the complexes is unknown. The frictional ratio used for the complex in these simulations was the one determined experimentally for the free interactants (1.4) but they are likely to be between 1.4 and the one expected for a globular complex (1.25). In both cases, the theoretical sedimentation coefficient difference of the 1:1 and 2:1 complex is 0.8 S and 0.9 S. So the peak shape observations made from these simulations with a frictional ratio of 1.4 would be the same for the other extreme case of a spherical complex.

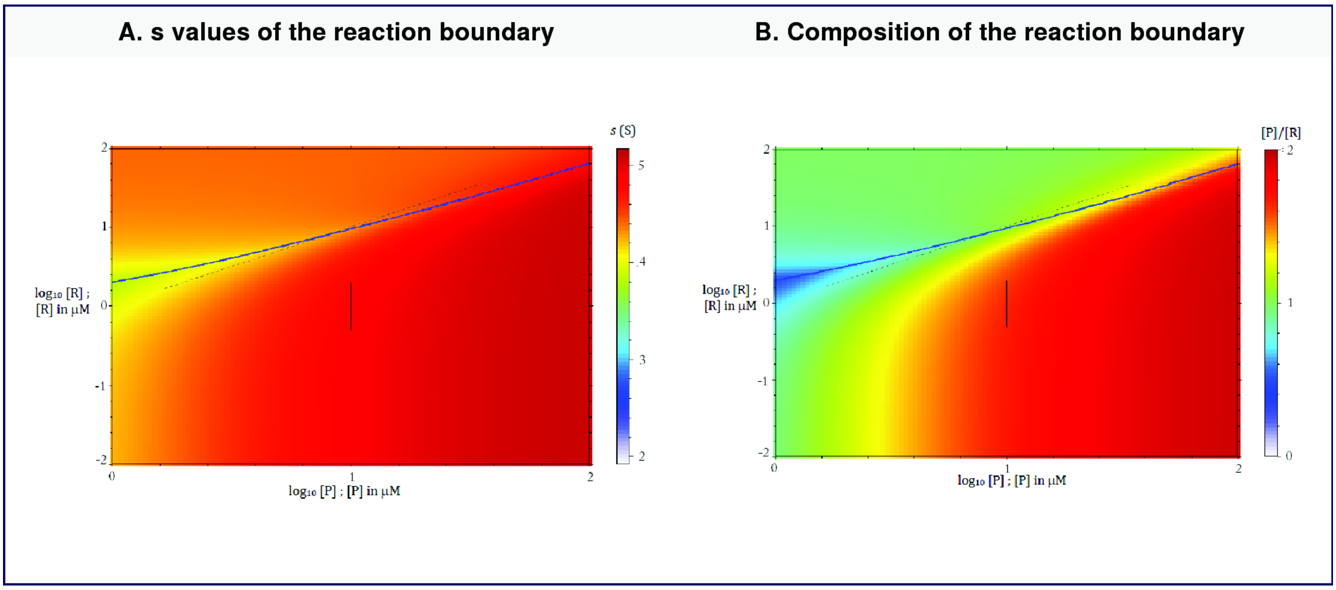
18.2.2 Simulation of the properties of sedimentation profiles using the effective particle explorer of Sedphat

Table 18.6: Summary of the fits of the simulated datasets with Lamm equation of non-interacting species analysis.

	Slow component			**Fast component**			
Simulation	s_{20w}	c	M_b (kDa)		s_{20w}	c	M_b (kDa)
1	1.92	0.11	5.63	4.89	0.14	18.55	
2	1.92	0.10	5.73	4.87	0.29	18.12	
3	1.92	0.10	5.53	4.71	0.29	16.23	
4	1.90	0.08	5.25	4.83	0.58	17.57	
5	1.92	0.11	5.63	4.90	0.14	17.84	
6	1.92	0.10	5.73	4.88	0.29	17.87	
7	1.93	0.10	5.35	4.72	0.29	15.55	
8	1.92	0.08	5.55	4.83	0.58	17.34	
9	1.92	0.11	5.79	4.92	0.15	15.10	
10	1.95	0.11	5.20	4.92	0.28	15.48	
11	1.98	0.12	4.98	4.79	0.27	13.22	
12	1.97	0.09	5.89	4.87	0.57	15.06	
13	1.96	0.10	4.64	4.44	0.56	14.25	
14	1.92	0.10	6.03	4.42	0.56	14.69	
15	1.87	0.07	9.18	5.17	0.59	20.02	

18.2.2.1 Description of the simulations

This part used a routine described by Schuck and colleagues (303), which is available within the Sedphat software ((10),v10.40 June 2012). The results are presented in **fig. 18.4** in the condition of Simulation 2 above (i.e. a simulation performed under a fast equilibrium with the equilibrium constants determined from ITC). The phase transition line (blue line) separates the concentration conditions for which the slow reaction boundary has the sedimentation coefficient of isolated R or P (upper and lower part of the figures, respectively). The black line represents the concentration condition we explored in our experimental AUC data. The color scale of the s-value of the reaction boundary plot in **fig. 18.4** A. gives s-values from 1.92 S (white) to 5.18 S (red). The color scale of the composition of the reaction boundary plot in **fig. 18.4** gives the molar ratio of TRBP-D1D2 over pre-miR-155 ([P]/[R]) from 0 (white) to 2 (red).


A. s values of the reaction boundary

B. Composition of the reaction boundary

Figure 18.4: Prediction of the *s*-value of the reaction boundary (**A.**) and composition of the reaction boundary (**B.**) using the effective particle theory implemented in Sedphat (10).

18.2.2.2 Results and conclusion

The experimental results are in agreement with this description, with the slow boundary at the *s*-value of the protein, the *s*-value and composition (evaluated experimentally from the ratio of A_{280nm}, A_{260nm} and interference signals) of the fast boundary are close to the value expected for the largest complex composed of two protein and one RNA molecule. The three experimental concentration conditions are similar in term of fast boundary *s*-value and composition.

18.2.3 Conclusion of the crossed svAUC and ITC approach

Thus under the assumption of a fast exchanging complex on the svAUC time scale, the result of these simulations match very well the experimental ones ; they are also good estimate of the sedimentation according Shucks and college in condition of intermediate exchange on sedimentation time scale. Perfect demonstration of the 1:1 and 2:1 complexes with this kind of simulation would require to repeat them with the other possible larger complexes but this would require to know their dissociation constants and obviously some conditions are expected to be also compatible with the experimental results presented here. Thus, the conclusion is that the high similarity between the data simulated with ITC knowledge and svAUC experimental data does not invalidate the conclusion that the 1:1 and 2:1 protein/RNA complexes are the essential complexes that can be formed in the solution conditions of this study. This compatibility validation between ITC and svAUC results was the main goal of these simulations.

18.3 ITC and svAUC combined conclusion: macroscopic model of TRBP interaction with microRNA 155 precursor

Although a presence of a small fraction of a complex of protein/RNA molar ratio of 2 and a stoichiometry higher than 2 protein: 1 RNA might exists, the combination of the ITC and svAUC data enable to conclude to the existence in solution of the following macroscopic models **fig. 18.5** for TRBP-D1D2.

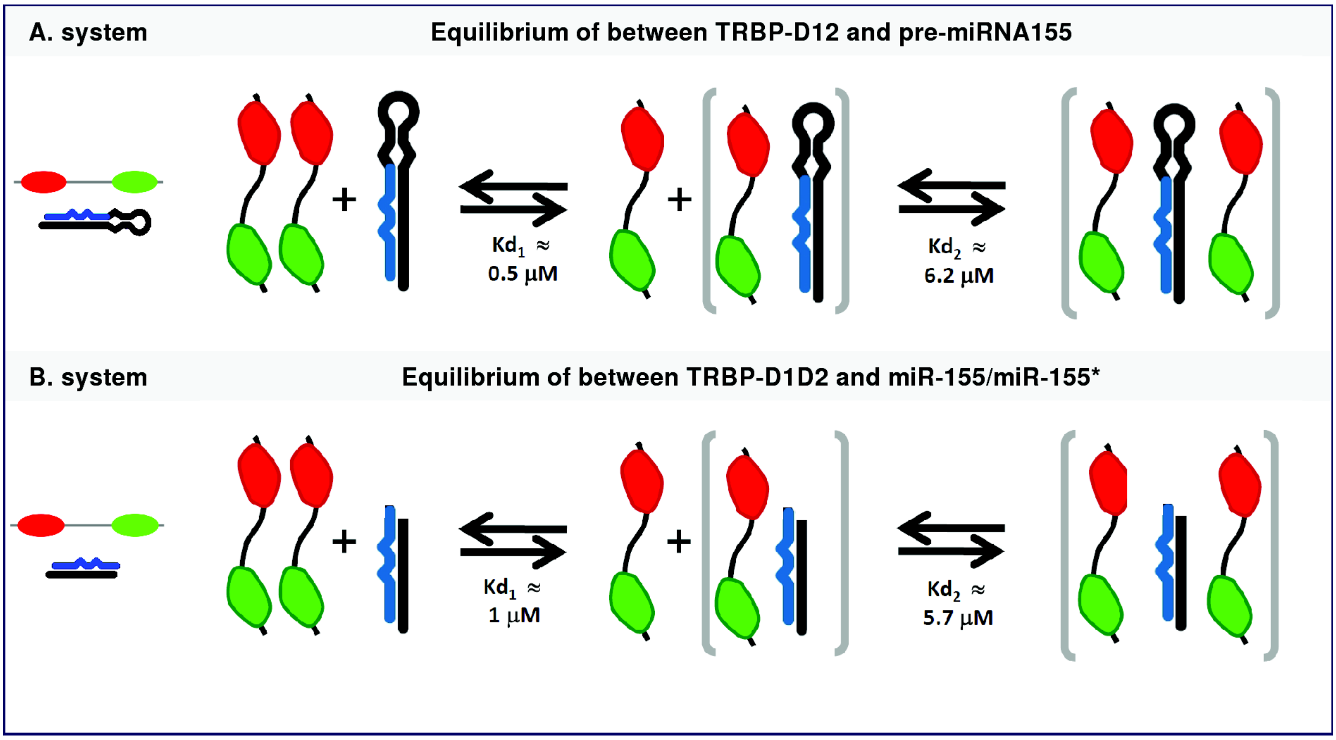


Figure 18.5: Illustration of TRBP-D1D2 and pre-miR-155 (**A.**) or miR-155/miR-155* (**B.**) macroscopic model from combined ITC and svAUC analysis.

For TRBP-D1D2, we can note that the first binding event has almost the same apparent affinity with the two precursors. The apparent affinity of the second binding event is slightly lower for the shorter precursor. Before starting the interpretation, I want to remind here that the shorter precursor engineered here is almost the equivalent of the product Dicer would lead from the cleavage of the longer one apart from the pair of GC nucleotides required for the *in vitro* production of this RNA. As a result the two RNA have a common stem, differs at the top of the stem by 4 nucleotides and one of them has an additional loop. Strictly, the difference is not just the presence/absence of a loop. Whatever the potential effect of this small sequence change, the similarity in the binding constants shows that the apical loop of pre-miR-155 is not required for the binding of two TRBP-D1D2. This does not mean that TRBP does not interact with it.

Multiple hypotheses can be proposed to explain the high similarity between the affinity of

the RNA binding region of TRBP and the two successive microRNA precursors of miR-155 :
- The RNA binding region of TRBP does not interact with the loop, it is restricted to the double stranded parts of the stem. This loop is undetected by NMR experiments on the imino protons regions which suggests that it is single stranded.
- The RNA binding region of TRBP interacts with the loop as strongly as it does with the engineered apical extremity (GC) of the shorter precursor. This hypothesis appear much less likely than the first one but can't be excluded from the available data.
- Coincidence could also explain the similarity. The measured constants are macroscopic ones and despite the relation between the first one and the second one are the ones expected in the case of a small negative cooperativity, much more complicated mechanisms potentially not identical with both RNA could lead to the same result.

Chapter 19

Site specific characterization of the RNA binding region of TRBP with microRNA 155 precursors: two independent binding domains

Combination of ITC and svAUC experiments have enabled to established a macroscopic model of the interaction between TRBP-D1D2 and pre-miR-155 and miR-155/miR-155*. Despite the RNA binding ability of each dsRBDs for pre-miR-155 has been established by ITC, the exact part of TRBP-D1D2 implicated in the interaction with miRNA precursors was yet to be established. This included the role of each dsRBDs in the two kind of complexes observed, the potential interaction between them on the RNA and the behavior of the interdomain region which was never reported even for siRNAs. Additionally, potential different binding behavior of TRBP on miRNA precursors before or after their cleavage by the enzyme Dicer (pre-miR-155 $\xrightarrow[\text{Dicer}]{\text{cleavage}}$ miR-155/miR-155*) was yet to explore. In particular, the effects of the presence of the loop were not reported in the literature. The macroscopic models suggested very similar affinities of TRBP-D1D2 for both precursors of miR-155. As both differences and similarity of the recognition of these RNA by TRBP would be an interesting result, a site specific comparison of the RNA binding region of TRBP to both precursors was performed. Since the macroscopic model of the interaction enable to choose the condition under which one of the two complexes is predominant, NMR is particularly well suited to answer the aforementioned objectives. Thus this chapter aims to determine the RNA binding surface of the miRNA binding region of TRBP (TRBP-D1D2) on pre-miR-155 and miR-155/miR-155*.

19.1 NMR assignment of the backbone of TRBP-D1D2, TRBP-D1 and TRBP-D2 on miR-155 precursors

In order to get insight into the behavior of the first two dsRBDs of TRBP in the tandem constructs, both TRBP-D1D2 and TRBP-D1 & TRBP-D2 were assigned in complex with pre-miR-155. miR-155/miR-155*, with lower yields than pre-miR-155 was used only for the interaction with TRBP-D1D2. As the study of the {TRBP / miRNA precursors} complexes by NMR was performed by looking at the TRBP signal, saturation of the protein required to add an excess of

RNA. Additionally, as the molecular weight of the {2 TRBP-D1D2: 1 pre-miR-155} is larger than the {1TRBP-D1D2: 1 pre-miR-155} complex (64.8 kDa *vs.* 42.2 kDa respectively) and because of exchange between free and bound partners in these systems, it was preferable to have only the 1:1 complex with the standard NMR experiments used (faster rotational correlation time and thus longer relaxation time for the smaller complex). The estimations of the macroscopic binding constants were used to design titration experiments by adding miR-155 precursors on solutions of TRBP constructs. The titrations starts from a state with a large proportion of large protein-RNA complex to a state were the equilibrium is displaced towards an excess of 1:1 protein:RNA complex **fig. 19.1**. RNA-bound states of the different TRBP constructs were analyzed using 2D ^{1}H-^{15}N BEST-TROSY experiments at various protein concentrations and protein/RNA ratios on [U-^{2}H,U-^{15}N] labeled samples with unlabeled pre-miR-155 and miR-155/miR-155*. The parameters of the different titration points reported here are given in **tab. 19.1** and **tab. 19.2**. The time required to collect individual titration points depended on the protein concentration and the signal-to-noise ratio (S/N). The experiments were recorded and analyzed as explained in Chapter 12. A titration with TRBP-D1D2 was performed with a [U-^{2}H,U-^{15}N,U-^{13}C] labeled sample up to a 2.3 excess of unlabeled pre-miR-155. At the end of the titration a sample with an RNA:protein ratio over 2.3 and a TRBP-D1D2 concentration of 230 μM was used to assign some (^{1}H,^{13}C and ^{15}N) backbone resonance assignments with a combination of 3D BEST HNCO and HNCA triple resonance NMR experiments. Partly because of the limited protein and RNA yields and because of the intermediate exchange of the interaction in the backbone NH planes, the TROSY signals of the complexes were weak. As a consequence, the common assignment strategy using sequential 3D experiments could not be used. Because of the moderate chemical shift change observed for most of the residues, the assignment was performed using the assignment of the apo forms as a basis. Using an overlay of the free and bound state HSQC spectra, from each free peak, the titrations enabled to reduce the number of possible candidates for the bound peaks. Then the comparison of the CO and inter/intra-residual CA, when present, enabled to complete the identification. The latter chemical shifts were very similar in the bound and free forms. Titration experiments were used for the assignment of the backbone amide resonances of TRBP-D1 and TRBP-D2 interacting with pre-miR-155 and for TRBP-D1D2 interacting with miR-155/miR-155*. Both titrations and the 3D HNCA and HNCO experiments were required for the assignment of the NH backbone of TRBP-D1D2. For a few residues, the titrations performed with the single domains were used to resolve the ambiguity on crowded titration signal in TRBP-D1D2 TROSYs. An additional small titration was performed by monitoring the imino proton spectrum of pre-miR-155 with injections of TRBP-D1D2 (**tab. 19.2**).

Table 19.1: RNA:protein ratios and expected total concentrations of the different titration points of TRBP-D1D2 by either pre-miR-155 or miR-155/miR-155*. The initial concentration of TRBP-D1D2 for both titration 500 μM in 70 μl, titrated by a solution of 2 mM pre-miR-155 or 970 μM of miR-155/miR-155*. C(X) indicate the loaded concentration of the species X.

	TRBP-D1D2 titration by pre-miR-155		**TRBP-D1D2 titration by miR-155/miR-155***	
$\frac{C(RNA)}{C(D1D2)}$	C(D1D2) (μM)	C(RNA) (μM)	C(D1D2) (μM)	C(RNA) (μM)
0	500	0	500	0
0.1	486	56	474	51
0.2	473	108	452	94
0.3	461	158	432	132
0.5	443	228	398	198
0.75	422	313	361	270
1.0	402	391	333	323
1.25	381	477	304	380
1.5	364	543	280	427
1.75	348	608	262	461
2.0	333	667	246	492
2.3	318	727	233	517

Table 19.2: RNA:protein ratios and expected loaded concentrations of different titration points of TRBP-D1 and TRBP-D2 by pre-miR-155. The initial concentration of TRBP-D1 and TRBP-D2 are 400 μM and 575 μM respectively with an initial volume of 55 μM in placed in 1.8 inner diameter Shigemi tubes. Additional points were recorded for TRBP-D1 at protein:RNA ratio of 0.2, 0.22, 0.29, 0.33 and 0.4. C(X) indicate the loaded concentration of the species X.

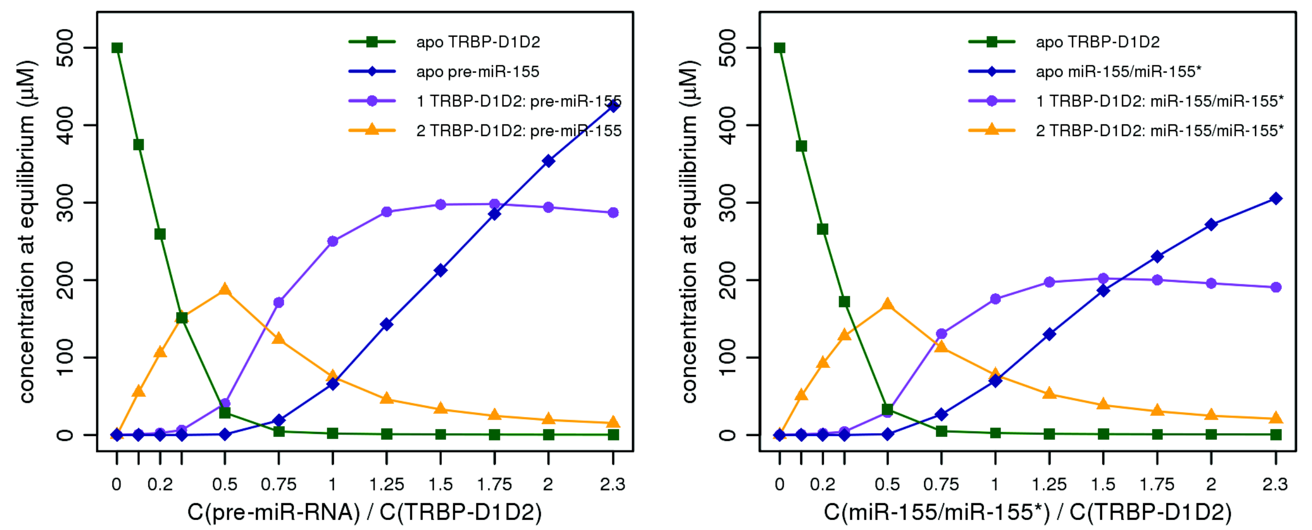
	TRBP-D1 titration by pre-miR-155		**TRBP-D2 titration by miR-155/miR-155***	
RNA: protein ratio	C(TRBP-D1) (μM)	C(RNA) (μM)	C(TRBP-D2) (μM)	C(RNA) (μM)
0	400	0	575	0
0.1	382	38	549	53
0.5	324	162	524	103
0.75	296	222	465	224
1.0	272	273	425	305
1.5	237	347	390	376

Table 19.3: TRBP-D1D2:pre-miR-155 ratios and expected loaded concentrations of different titration points of TRBP-D1 and TRBP-D2 by pre-miR-155. C(X) indicate the loaded concentration of the species X.

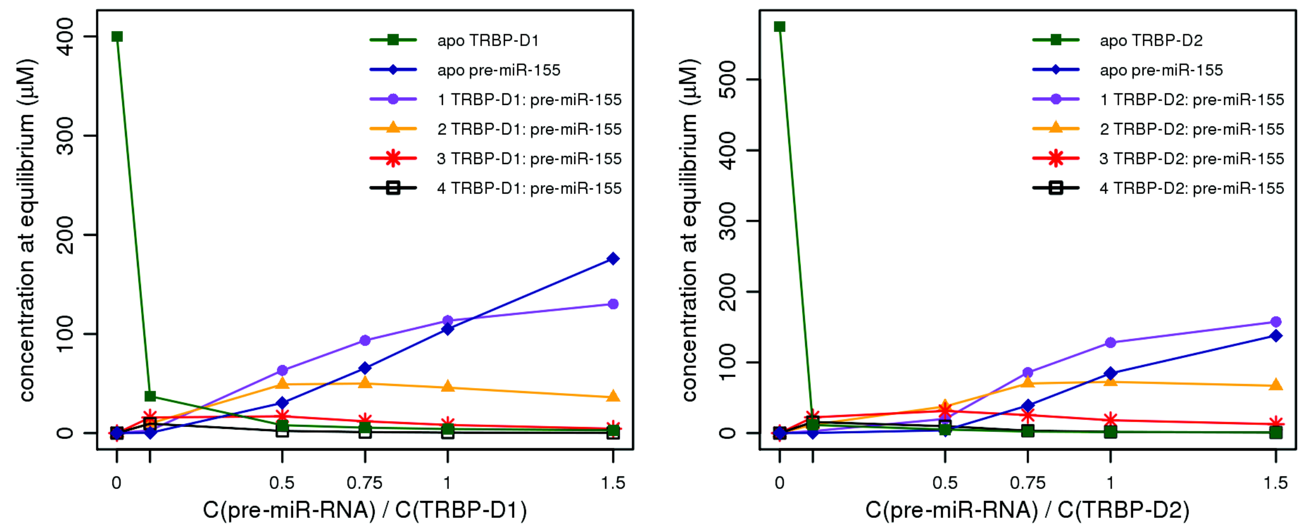
TRBP-D1D2: RNA ratio	C(RNA) (μM)	C(TRBP-D2) (μM)
0	200	0
0.1	191	21
0.5	165	86
1	140	146
2.0	108	222

19.2 Results

Spectra corresponding to the interaction between TRBP constructs and miR-155 precursors are given in **fig. 19.2**. Quantification of the combined proton and nitrogen backbone chemical shift are given in **fig. 19.3**.


A. TRBP-D1D2 titration by pre-miR-155

B. TRBP-D1D2 titration by miR-155/miR-155*


C. TRBP-D1 titration by pre-miR-155

D. TRBP-D2 titration by pre-miR-155

Figure 19.1: Prediction of the evolution of the concentration of the different protein/RNA complexes and single species during the NMR titrations. The calculation used the macroscopic models for TRBP-D1D2 with pre-miR-155 and miR-155/miR-155* derived without microscopic assumptions, and the rough constants estimates for TRBP-D1 and TRBP-D2 with pre-miR-155 derived under the hypothesis of identical and independent binding sites of pre-miR-155

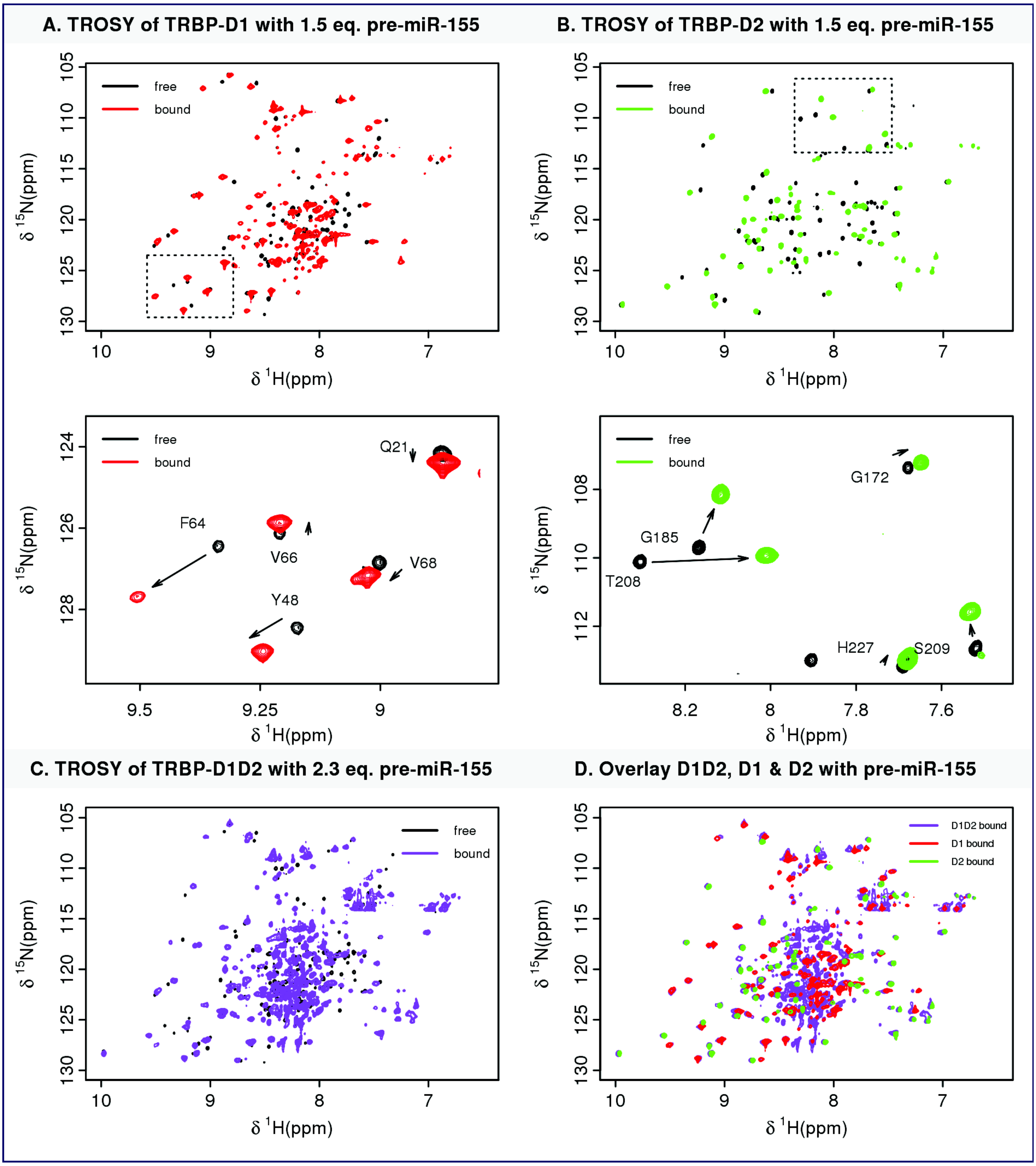


Figure 19.2: (^{1}H,^{15}N) TROSY-HSQC spectrum of (U-^{2}H$_2$O-^{15}N)-labeled TRBP-D1 (A.), TRBP-D2 (B.) and TRBP-D1D2 (C.) in the free state or interacting with pre-miR-155 with 1.5, 1.5 and 2.3 RNA equivalent respectively. The spectra were recorded at 800 MHz at 25°C in [23.5 mM potassium phosphate pH 6.5, 100 mM KCl, 10 mM $MgCl_2$ and 5 mM β-mercaptoethanol, 2 mM TCEP]. The panels of the second row show an expansion of the region of the full spectra of the first row contained in the dashed box ; assignments of a few peaks are given as example. Panel D. is an overlay of the TROSY spectra of the bound state of TRBP-D1 (A.), TRBP-D2 (B.) and TRBP-D1D2 (C.) interacting with pre-miR-155 with 1.5, 1.5 and 2.3 RNA equivalent.

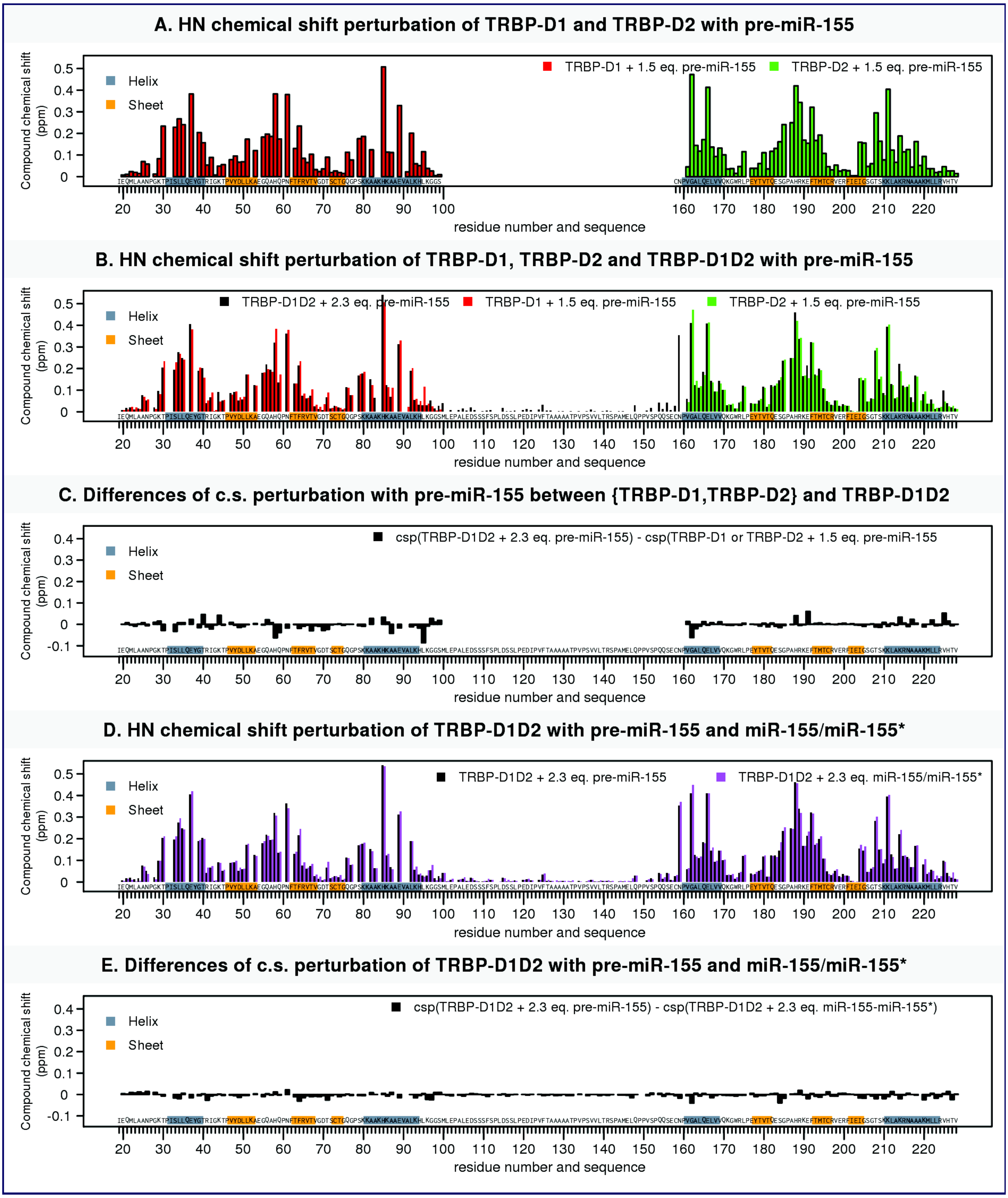


Figure 19.3: Comparison of backbone amide chemical shift (c.s.) perturbation of TRBP-D1D2, TRBP-D1 and TRBP-D2 upon interaction with pre-miR-155 or miR-155/miR-155*. The hydrogen and nitrogen chemical shift are calculated with the compound chemical shift formula given in Chapter 12. Secondary elements are taken from the secondary structure prediction of free TRBP-D1D2 performed in this work from NMR chemical shift in solution. The vertical scale for the chemical shift perturbation and differences of chemical shifts perturbations is the same for all plots.

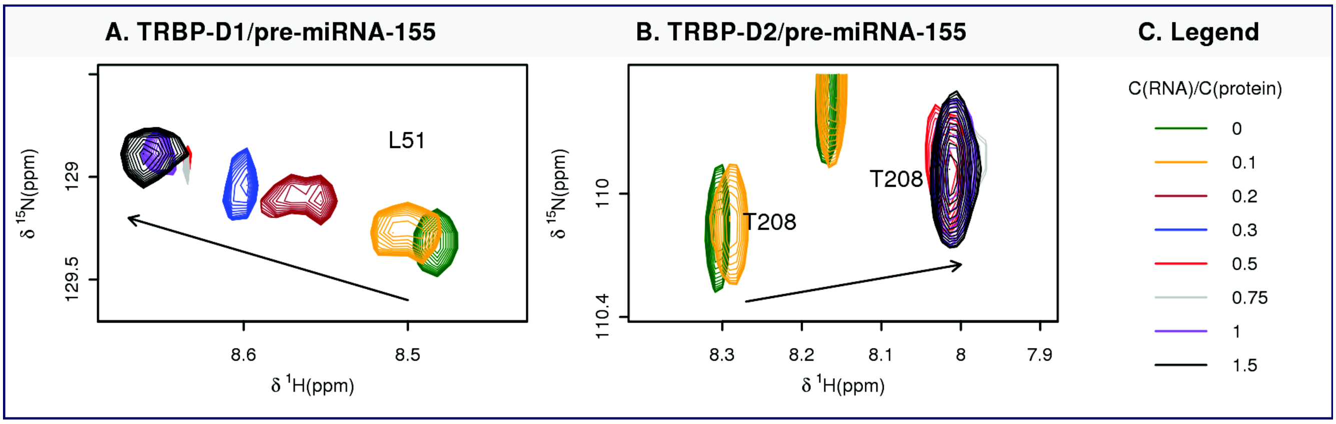


Figure 19.4: Typical examples of peak behavior in titrations of TRBP-D1 and TRBP-D2 with pre-miR-155.

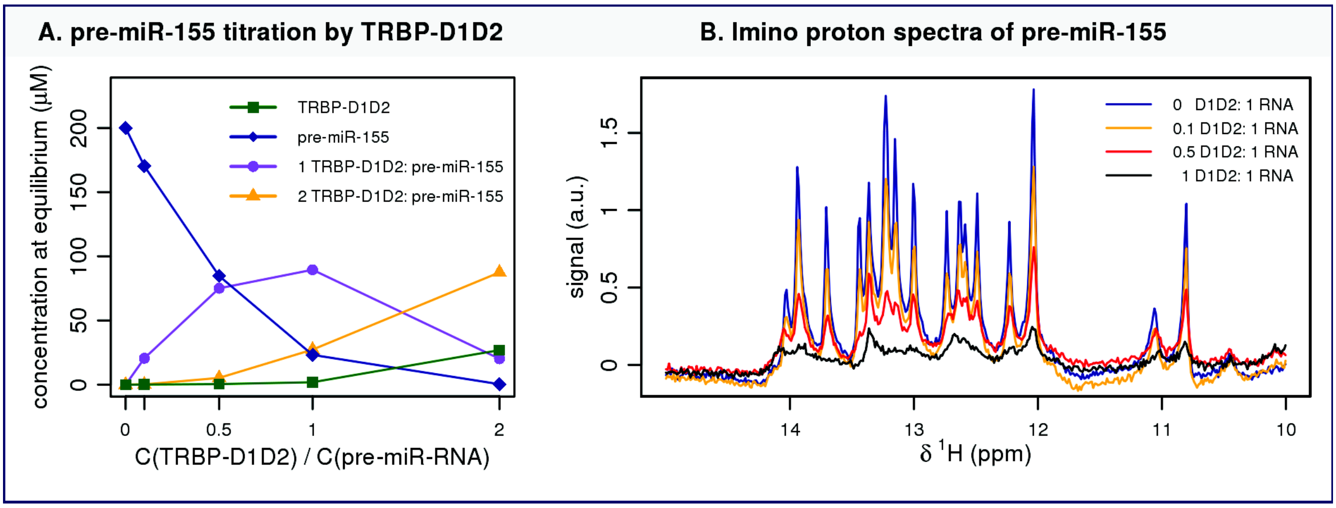


Figure 19.5: Titration of pre-miR-155 by TRBP-D1D2 monitored by recording 1D imino protein spectra (SOFAST-HMQC) at 25°C. **A.** Prediction of the evolution of the concentrations during the titration using the macroscopic model established in this study. **B.** Evolution of the imino proton spectrum of pre-miR-155 during the titration. All the experiments were recorded with the same parameters and total time of acquisition.

Information on {TRBP-D1D2/miR-155 precursors} complexes For TRBP-D1^{F}D2, TRBP-D1^F and TRBP-D2 interacting with pre-miR-155, 93.2%, 84% and 97% respectively of non proline backbone amide resonances were assigned. 62% of non proline CO and 66% of non proline Cα were assigned for TRBP-D1^{F}D2 interacting with pre-miR-155 and 93.2% of non proline backbone amide resonances for TRBP-D1^{F}D2 with miR-155/miR-155*. Most of the peaks of the proteins constructs have different position and diminished intensities upon the mixture of the RNA solution and the protein solution. This implies a change of the chemical environment of the amide protons. As both macromolecules are in the same buffer, these changes correspond to the interaction between the TRBP constructs and miR-155 precursors. Titrations were characterized by a strong diminution of the signal intensities characteristic of NMR intermediate exchange between free and bounds species. Two different regimes for TRBP-D1 and TRBP-D2 were observed and are illustrated in **fig. 19.4**. The exchange for TRBP-D2 is globally closer to slow exchange than the one of TRBP-D1, closer to fast exchange. These differences reflect kinetic differences that are consistent with the higher affinity of TRBP-D2 compared to TRBP-D1. When detected, the peaks displacements were almost linear. No significant chemical shifts are observed for the unfolded form of the first dsRBD of TRBP in either TRBP-D1 or TRBP-D1D2 ; however the relative intensity of this form over the folded one diminish as RNA is added to the protein solution, which may corresponds to an expected displacement of the F = U equilibrium towards the RNA interacting form. The chemical shift perturbations are given for the end of the titrations. Observation of the imino protons signals of pre-miR-155 with TRBP-D1D2 **fig. 19.5** mainly show the spectrum of the free RNA in solution according to the prediction of complexes concentration. Providing the free RNA is not in NMR slow exchange with the bound form the lack of peak shift suggests no major structural change on the RNA in the complex. TROSY experiments are more informative ; according to the predictions of the complex abundance in **fig. 19.1** the system for which chemical perturbation are given correspond to complexes with 1 TRBP-D1D2 on pre-miR-155 or on miR-155/miR-155* and one or two TRBP-D1 or TRBP-D2 on pre-miR-155. Using this information, the backbone chemical shift comparisons and assuming the backbone amide chemical shifts reflects the behavior of the protein/RNA interaction surface, the following conclusion can be driven :

- Both dsRBDs of TRBP-D1D2 are involved in the interaction with one molecule of pre-miR155 or miR-155/miR-155*. The interdomain region is not directly involved in RNA recognition. Absence of significant shifts and high intensities of interdomain region peaks suggests by comparison to free TRBP-D1D2 that this region remains flexible in the TRBP-D1D2/RNA complex. Thus this region can be qualified as a tether.
- Amplitudes of chemical shifts perturbation in TRBP-D1D2, TRBP-D1 and TRBP-D2 interacting with pre-miR-155 are almost the same. This implies that the chemical environment of the two dsRBDs in the tandem construct TRBP-D1D2 is almost the same than when they are allowed to interact freely with pre-miR-155. Thus, both dsRBDs of TRBP-D1D2 interact independently on the RNA in a system where the {1 TRBP-D1D2: 1 pre-miR-155} is predominant. Three regions of the dsRBDs have large chemical shift perturbation. These regions correspond to the three canonical RNA binding regions of dsRBDs presented in part I. So in addition to the independent binding, the first two dsRBDs of TRBP interact with the canonical dsRBD RNA binding surface. Similar secondary structure chemical shifts are compatible with no large structural changes during the binding and in particular not change of secondary structure. A similar conclusion was deduced from the study of the crystallized form of a complex between the second domain of TRBP and a small RNA duplex (290). So it is possible and useful to plot the chemical shift perturbation on the available free structures of the first two dsRBDs of TRBP (**fig. 19.6**).
- Amplitudes of chemical shifts perturbation in TRBP-D1D2 interacting with pre-miR-155 or

with miR-155/miR-155* are almost the same. This imply that the RNA binding surface of TRBP-D1D2 with either pre-miR-155 or miR-155/miR-155* in a system where the {1 TRBP-D1D2: 1 RNA} is essentially the same. This observation is consistent with the similar affinities for these complexes derived from ITC experiments.

Importantly, despite the NMR results give the evidence that 2 dsRBDs of TRBP can bind the same precursor molecule, the present results corresponds to an average information on an ensemble of molecules and cannot exclude the possibility that the dsRBDs are not exactly bound at the same time.

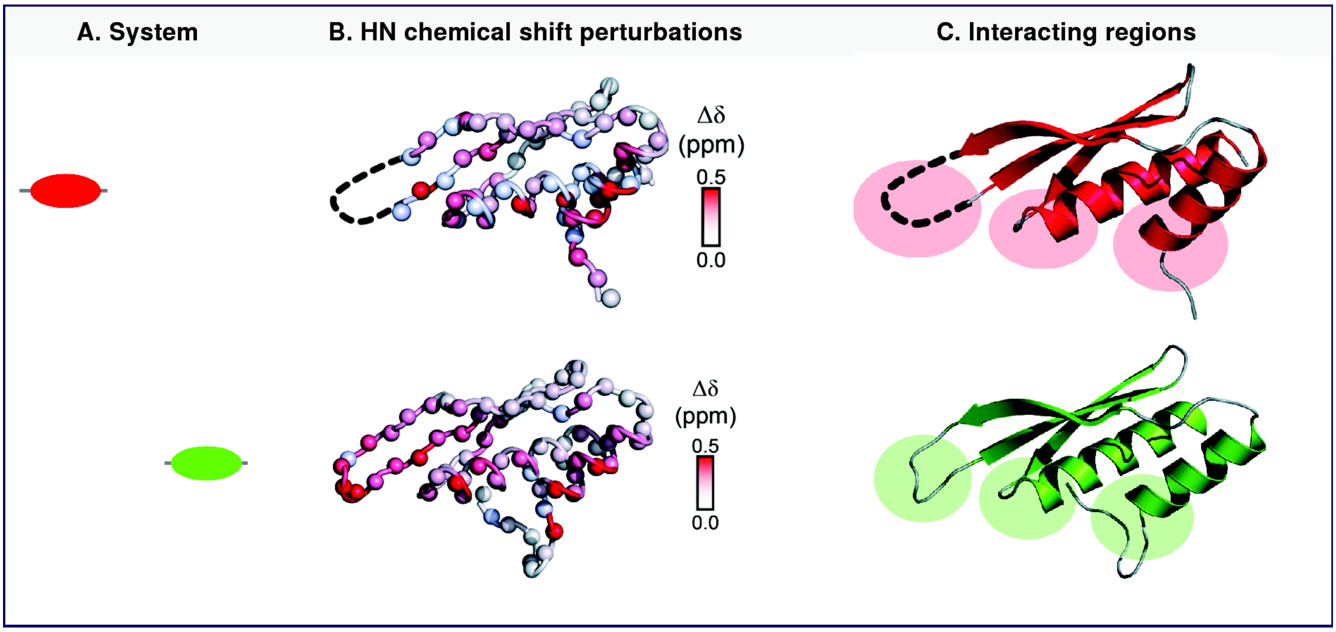


Figure 19.6: Representation of chemical shift pertubation on the structures in the apo state of the first two dsRBDs of TRBP (PDBid: 3LLH, 2CPN).

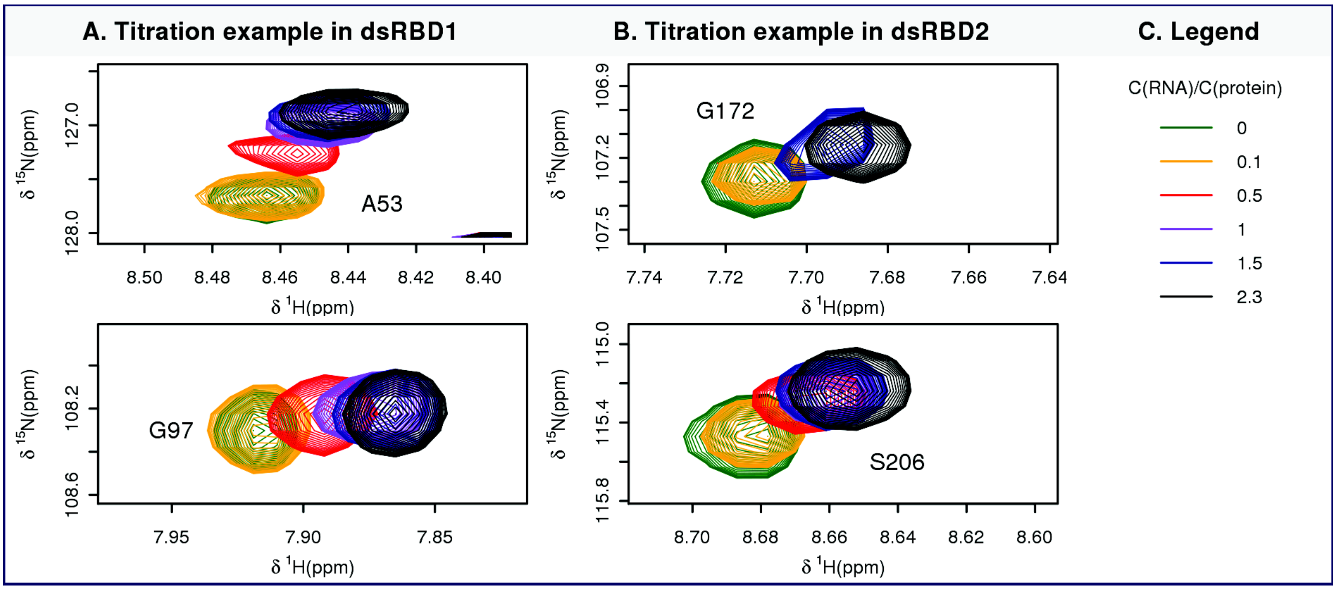


Figure 19.7: Typical examples of peak behavior in titrations of TRBP-D1D2 with pre-miR-155.

NMR derived information on {(TRBP-D1D2)$_2$ / miR-155 precursors} complexes Previous conclusions were derived from experiments based on the observation of TRBP-D1D2 signal with an excess of RNA, shifting the equilibrium towards a 1:1 protein:RNA complex. With the available experiments, structural information of the complex with two TRBP-D1D2 on miRNA precursors can only be obtained from the beginning of the titrations where this complex is predominant. However, there is also an excess of free protein at the beginning of these titration so the signal reflect largely this species before an RNA:protein ratio of 0.4. Titration example of typical behavior of peaks from dsRBD1 and dsRBD2 in TRBP-D1D2 are given in **fig. 19.7**. The signal to noise of the spectra in the first points of the titration and the intermediate exchange did not allowed to perform a systematic study for all peaks and especially for the ones with the largest chemical shift changes. If significant non linearity in the peak displacement would be detected, including peaks at a stage were the 2:1 protein:RNA complex is highly predominant ($\frac{C(\text{RNA})}{C(\text{TRBP-D1D2})} = 0.5$), this would indicate a different structure of the complex with two TRBP-D1D2 on one miR-155 precursor. The rather linear displacement does not enable to make conclusion on the properties of this complex from the detectable peaks in these experiments.

Chapter 20

Conclusion and discussion of the main results

Much of the information concerning the roles of TRBP in small silencing RNA synthesis (miRNAs and siRNAs) has been obtained from studies with siRNAs whereas miRNAs are a more important small RNA class in humans. The 3D structures of the first two dsRBDs of TRBP have been published (290) as well as the structure of TRBP-dsRBD2 in complex with a short 10 bp dsRNA (294) but without information on the actual binding sites, binding behavior, stoichiometry with microRNA precursors, behavior of the free partners, etc Little biophysical or structural analysis of how multiple domain constructs of TRBP interact with miRNA precursors have been previously reported. The data presented here represent the first, in depth biophysical analysis of the interaction between TRBP and human miRNA precursors which have an imperfect dsRNA stem. The main conclusion of the TRBP/miR-155 precursors interaction found is this thesis are summarized in the following paragraphs and compared with other similar systems reported in the literature, in the light of the first part of the thesis.

Importantly, it has to be underlined here that the conclusions proposed from this study are valid under the *in vitro* conditions used. The *in vivo* behavior in the crowded in cell environment might be different. It is however expected that this study on isolated macromolecules reveals some of their fundamental properties that should contribute to what happen actually *in vivo*.

RNA binding sites are restricted to the dsRBDs The RNA binding sites in the RNA binding region of TRBP are restricted to the two dsRBDs and do not involve the interdomain region. This region is flexible in the free and bound form of the protein. The existence of such a flexible linker has been previously demonstrated in PKR, ADAR2 and HYL1 where a long unstructured linker connects the two dsRBDs.

The two first domains of TRBP behave independently in the free and bound form on the protein As presented in part I, dsRBDs are part of modular RNA-binding proteins, either in multiple copies (mainly tandems) or in combination with other functional domains such as RNase III, RNA Helicases, PAZ domains, etc. All solution studies in dsRBD tandems, using liquid state NMR spectroscopy, have shown that the dsRBDs were independently folded, with no intramolecular interaction between these domains in the apo form: first two dsRBDs of HYL1 (231), the two dsRBDs of human PKR (203) and human ADAR2 (256). The two dsRBDs of human ADAR2 bind to distinct regions of the dsRNA stem without inter-domain contact or contribution from the inter-domain linker (257; 256).

The work performed is this thesis has shown that the first two dsRBDs of the RNA binding region of TRBP (TRBP-D1D2) are also independent in the apo form and in the RNA holo form to either one pre-miR-155 or one miR-155/miR-155*. SEC-MALLS and svAUC demonstrated the absence of detectable intermolecular interaction between the dsRBDs in the apo form and NMR chemical shift perturbation studies on TRBP-D1D2, TRBP-D1 and TRBP-D2 enabled to show the independency of the dsRBDs in the apo form and the independency of their RNA binding surface in the 1:1 protein:RNA complex.

The two first domains of TRBP binds to the same RNA and up to two copies can bind on miR-155 precursors A second important feature of the interaction between the RNA binding region of TRBP and cytoplasmic miR-155 precursors is the stoichiometry. ITC combined with svAUC experiments has enabled to propose macroscopic models for the interaction between TRBP-D1D2 and pre-miR-155 or miR-155/miR-155*. For both precursors, the models showed the presence of 1:1 and 2:1 TRBP-D1D2:RNA complexes. Pre-miR-155 was shown to be mainly monomeric using both SEC-MALLS and svAUC and its secondary structure is compatible with the predicted miRNA precursor hairpin. Knowledge of the state of the precursor is of high importance to have an accurate understanding of the stoichiometry in complexes involving dsRBDs since the longer the dsRNA, the more dsRBDs are expected to bind. The stoichiometry of 2 is the first reported for TRBP with miRNA precursors. A study of the interaction between TRBP and a siRNA, performed during the course of the present study, has reported by ITC that two TRBP-D1D2 (similar construct) and that full length TRBP could bind to a 19 base pairs siRNA duplex with perfect base pairing but without structural information on the complexes. This result is consistent with the observation made here with miR-155 precursors. HYL1, a dsRBD containing cofactor of *Arabidopsis thaliana* DCL1 has been shown to form 2:1 protein:RNA complexes with dsRNA sequences of suitable length (294).

In contrast to the 1:1 protein:RNA complex, no structural information was obtained here for the complex with 2 TRBP-D1D2 and one miRNA precursor by NMR. The macroscopic model could not exclude preferential binding of only one of the two dsRBDs of TRBP-D1D2 in this complex. ITC experiments with the isolated dsRBDs (TRBP-D1 and TRBP-D2) of TRBP also indicate their ability for form a complex containing 4 dsRBDs on one pre-miR-155. This observation (4 independent dsRBDs interacting with one precursor) suggest that each dsRBD in the [2 TRBP-D1D2]/pre-miR-155 complex could bind the miRNA precursor.

The functional stoichiometry of the Dicer/TRBP complex has not yet been determined (156). A study reports that a tertiary complex containing Dicer, TRBP and PACT can be formed which implies that two TRBP-like proteins can interact simultaneously with Dicer in the presence of pre-miRNAs (148). This corresponds to 8 dsRBDs in a complex, some of them being potentially not able to bind dsRNA. Thus, the high occupancy of dsRBDs on a relatively short sequence of dsRNA, as seen here with TRBP and two miRNAs precursor, is potentially functionally relevant. Whether the 2:1 ratio observed here is a standard feature of the complexes formed between TRBP and miRNA precursors or whether the stoichiometry is determined by features in the target miRNA precursor, such as the length or percentage of canonical dsRNA, remains to be determined. Additionally, the possibility of a preferential position of the dsRBDs on the precursors in the 1:1 and 2:1 protein:RNA complexes could be worth to investigate.

The behavior of TRBP-D1D2 is similar with both hairpin and duplex miR-155 precursors RNA experiments have shown that in the 1:1 TRBP-D1D2:RNA complex, the protein interaction surface was identical with pre-miR-155 and miR-155/miR-155*, which are the cytoplasmic precursors of miR-155 before and after Dicer cleavage respectively. The proposed macroscopic models also provided information on the affinity of the different binding events of TRBP-D1D2 for these two precursors. Analysis of ITC data yielded dissociation constants for the TRBP/pre-miR-155 complex in the low μM range, which is consistent with other investigations of dsRBD/dsRNA interactions (231; 294; 252; 57; 284). Importantly the dissociation constants are very similar for the two precursors, a comparison that was not studied before. A recent study on TRBP with similar construct and an additional 6 histidine tag (290) reported by ITC a much tighter binding in the sub nanomolar range with a 19 base pair perfect siRNA duplex. The affinity difference with the one estimated in the present study could come from the RNA used (imperfect base paired miRNA precursors here *vs.* siRNA perfect duplex, sequence effects) and/or an effect of the protein purification tag and/or an influence of the different buffer used (20 mM Hepes-NaOH buffer (pH 7.5) with 150 mM NaCl in their study). However, the moderate affinity of the RNA binding region on TRBP with miR-155 is compatible with the displacement of the domains by the enzyme Dicer or other proteins of the RISC loading complex that process the RNA.

Despite the independency of the RNA binding surfaces of each dsRBD in TRBP-D1D2, a complex network of cooperativity interaction with the miRNA precursors, especially in the 2 protein:1RNA complex cannot be excluded. It can however be noted that the relation between the first and second macroscopic binding constants corresponding to the successive binding of two TRBP-D1D2 on the precursors are compatible with a quasi independent/slight negative cooperativity interaction of these two proteins. The potential absence of positive cooperativity result would be consistent with earlier studies, which compared the RNA-binding properties of TRBP with *C. elegans* RDE-4 (214).

The very similar affinities of TRBP for pre-miR-155 and miR-155/miR-155* duplex is of interest. Whether the protein binding sites on the RNA and whether the loop is involved cannot be rigorously obtained from this information. However with the identical RNA binding surface for TRBP-D1D2 with pre-miR-155 and miR-155/miR-155* in the 1:1 complexes strongly suggest that TRBP-D1D2 interacts on the substrate and product of Dicer via the same RNA sites. In addition to its role in the Dicer complex, TRBP is a component of the RLC (35; 106; 182). The similar affinities of TRBP-D1D2 and protein binding surfaces suggest that TRBP may have *in vivo* an RNA binding activity upstream and downstream of the cleavage of miRNA precursors by Dicer, an ability that was not yet clear. Therefore TRBP may participate in the recognition of Dicer substrates as well as the transfer of the cleavage product to RISC via the RISC loading complex. To illustrate the main interaction results and to investigate the steric possibility of 2 TRBP-D1D2 (4 dsRBDs) on each miR-155 precursors with no interdomain contacts, a structural model of the complexes has been generated using Rosetta (49). It is combined with the macroscopic model in **fig. 20.1** The description of the generation of this model is given in the second publication made during this thesis (15). The model uses the TRBP-dsRBD2 construct bound to an RNA duplex (PDBid: 3ADL). Although no information are available on the relative position of each dsRBD in these complexes, the double stranded RNA surface might allow a 2 protein:1RNA complex without significant dsRBD contact. Experiments with deuterated and ^{15}N labeled RNA could advantageously enable to work in conditions with saturated 2:1 protein:RNA complex and try to further characterize the largest complex.

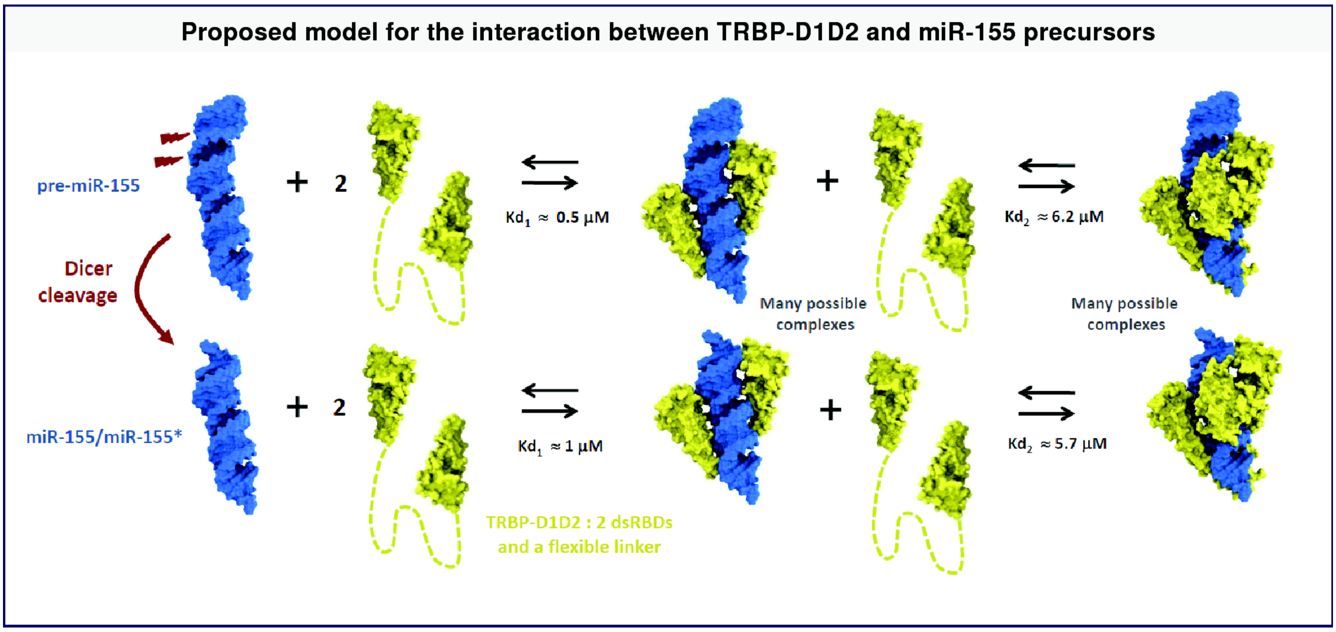


Figure 20.1: Possible model of the interaction between the RNA binding region of TRBP and miR-155 precursors using the macroscopic model established from SEC-MALLS, ITC and svAUC combined with the site specific interaction of the 1:1 complex performed by NMR and showing an independent behavior of the dsRBDs on the precursors.

The binding affinity and dynamics of the two first dsRBDs of TRBP are different Differences between the behavior of each dsRBD in TRBP-D1D2 were also detected in the present study. First, the affinities measured for the isolated dsRBDs indicate that TRBP-D2 interacts with pre-miR-155 about 4 times more strongly than TRBP-D1. The relative affinity is consistent with a study performed with similar isolated dsRBDs of TRBP, interacting with a 19 base pair siRNA duplex (290) in which the affinity of the second dsRBD is twice the one of the first one. As developed in part I, differences of affinity in dsRBDs in tandems are very common but the potential roles are unclear.

Another striking difference discovered in this thesis is the high dynamic difference between the two first dsRBDs of TRBP: the second domain appears to be stable on the sampled time scales whereas the first domain undergoes a full folding/unfolding equilibrium slightly shiftable towards the folded form in the presence of RNA. According to the author knowledge, this full folding/unfolding process is the first one reported for a human protein in conditions close to physiological ones in pH and temperature. The *in vivo* relevance of this interesting mechanism remains to be determined as well as the potential links with the affinity. The difference in dynamics between two dsRBDs domains has already been reported for the pair of dsRBDs in PKR (204), but evidences of folding-unfolding equilibrium. These two domains were also reported to have also different RNA binding affinities. The study proposes "a direct correlation between the conformational flexibility at a millisecond timescale and dsRNA binding activity". The proposed interpretation is that there is a concerted motion at the milliseconds-microseconds timescale in the first dsRBD with the higher RNA binding affinity generate multiple conformations required for the binding of different RNAs. The argument is however weak as the "correlation" observed for 2 individuals and 2 variables might just be coincidence. As they report, the binding sites of these two domains is known and the residues involved are different in the two domains. This chemical difference could also explain/contribute to the difference. Additionally, the authors have shown that the second domain, more rigid can interact with the third one, a property that could be suffice to explain the natural selection of a more stable second dsRBD, independently to the RNA affinity variable. Both the protein and the RNA structures & dynamic have an effect in this type of interaction so the deconvolution of the different processes is challenging. The dynamic range on the isolated protein side does not indicate the type of interaction mechanism that will happen with the RNA, which can range from conformation selection to induced fit types of recognition. A conclusion that can be made from the present study and the one on PKR dsRBDs is that in addition to the common affinity differences in dsRBDs tandems, significant dynamical behavior exists in two studies. The consequences of these differences on the RNA interaction are still very elusive. Only hypotheses can be formulated. It is possible that the higher number of conformations adopted by one domain allow it to recognize a large variety of substrates. Because these pairs of dsRBDs are usually involved in multiproteins complexes, it is also possible that the weaker affinity of one domain enables it to be displaced more easily than the other one by another RNA binding protein. A more detailed study of the dynamics of dsRBDs on their RNA target in a multiple systems in a competitive multi-protein environment seems required to further investigate in this direction.

Proposition of future research directions In the future, it would be interesting to test whether the properties found here for TRBP-D1D2 are influenced by the presence of the third domain of TRBP which may shift the equilibrium towards the 2:1 protein:RNA complex because of its homodimerization capability. Additionnaly the effects of the phosphorylation state of the protein and of the binding of the anticancer drug Enoxacin are interesting. Preliminary studies performed during this thesis have suggested that it was respectively not the case with phosphomutants and slightly the case with enoxacin.

The original behavior of the first domain of TRBP, undergoing a folding-unfolding exchange in a wide range of *in vitro* conditions merits further investigations, in particular the existence of a similar phenomenon in other dsRBDs and assessment of the its potential functional relevancy. In parallel to the work reported here, a collaborator of the host lab, found that a plant dsRBD in a dsRBD tandem was fully unfolded in the free state and able to fold upon addition of the miRNA precursor. The 1:1 protein:RNA complex obtained in this case is a mixture of a folded form bound to the RNA and of the apo unfolded form. I performed EXSY experiments on this complex which help the team to assign the RNA bound state and demonstrated the exchange (to be published). This dsRBD appears to be another interesting case in the dsRBDs family. Together with the behavior reported here for TRBP-D1, it might indicate a much larger heterogeneity in dsRBD conformations that what is suggested by the available structures.

It would be interesting to relate the properties found for TRBP-D1D2 with the recent report of a movement of TRBP on dsRNA (147) for which the first two dsRBDs are indispensable. This property of TRBP, also found in PACT, enable the Dicer-TRBP complex to diffuse on the dsRNA, which was not detected for Dicer alone with dsRNA. The authors propose a role for this process in the correct positioning of Dicer and on its cleavage rate. This process might be important for the processing of siRNA which originate from long dsRNA precursors. It would be interesting to study the influence of the linker on this process and the influence of the different properties of the two dsRBDs including the affinity difference and the folding/unfolding equilibrium in TRBP-D1. The use of chimeras with tandems of dsRBD2 or dsRBD1 could help in these investigations. Finally, the influence and potential conservation of TRBP-D1D2 properties in the context of larger complexes with the enzyme Dicer and within the RISC loading complex are obvious future research directions.

Part VI

Annexes

20.1 Résumés en francais

20.1.1 Résumé général

Les microARNs (miARNs) sont une classe de petits ARNs non codants qui régulent l'expression des gènes via un mécanisme d'interférence par ARN. Les microARNs humains sont produits par une série de réactions enzymatiques. En particulier, dans le cytoplasme le précurseur de miRNA (pre-miRNA) est reconnu et clivé par un complexe contenant l'enzyme RNase III Dicer et plusieurs cofacteurs protéiques. La protéine TRBP (HIV TAR RNA binding protein) est l'un de ces cofacteurs ; il augmente la stabilité du complexe, influe sur la cinétique, la position du clivage et a rôle potentiel dans la reconnaissance du substrat et dans le transfert du produit vers le complexe RISC (RNA-induced silencing complex) effecteur de l'interférence par ARN. TRBP est composé de 3 domaines de liaison aux ARN doubles brin (dsRBDs). La région d'interaction de TRBP avec les ARNs est composée des deux premiers dsRBDs liés par une région interdomaine non caractérisée. La présente étude est une caractérisation biophysique *in vitro* de la région d'interaction avec les ARNs de TRBP dans l'état apo de TRBP ou dans l'état lié avec chacun des deux précurseurs cytoplasmiques successifs du microARN oncogène miR-155 comprenant la tige boucle pre-miR-155 et le duplex miR-155/miR-155* résultat du clivage de pre-miR-155 par Dicer. L'étude montre que la région d'interaction de TRBP avec les ARNs est monomérique, est composée de deux dsRBDs de conformations indépendantes en solution et que la région interdomaine de 60 résidus est flexible. Le premier dsRBD, non caractérisé précédemment en solution est le siège d'un équilibre plié/déplié intégral dans une grande gamme de conditions physico-chimiques. Les deux premiers dsRBDs de TRBP peuvent interagir avec un même précurseur de microARN et deux régions d'interaction de TRBP avec les ARNs peuvent interagir avec un même precurseur. La région d'interaction de TRBP avec les ARNs interagit avec pre-miR-155 et le duplexe miR-155/miR-155* avec des affinités très similaires. Dans le complexe avec une région d'interaction de TRBP avec les ARNs liée à pre-miR-155 ou au duplex miR-155/miR-155*, aucun indice de contact entre les deux dsRBDs n'a été detécté et la protéine interagit avec les deux précurseurs par la même surface d'interaction. Les informations récoltées suggèrent notamment que TRBP peut jouer un rôle avant et après le clivage des pre-miARNs par Dicer, en particulier dans le complexe de chargement de RISC.

20.1.2 Résumé de la partie I

La partie I présente le contexte biologique dans lequel le travail de thèse a été effectué. Elle est constituée d'une revue critique de la littérature accompagnée d'analyses structurales. A partir de l'an 2000, l'identification progressive de petits ARNs régulateurs présents chez les plantes et les animaux a prouvé l'existence d'un nouveau mode général de régulation génique. Les petits ARNs impliqués, d'environ 20 à 30 nucléotides, sont responsables d'un mécanisme d'interférence par ARN dans lequel leur interaction avec un ARN messager engendre la dégradation ou le stockage de ce dernier. Six principales classes de petit ARNs régulateurs sont généralement reconnues. Les microARNs canoniques, les microARN mirtrons, les microARNs dérivés d'ARNs fonctionnels, les siARNs (si : short interfering) endogéniques et exogéniques et les piwiARNs. Le travail de thèse s'est focalisé sur la biogenèse des microARNs (miARNs) et siARNs dont la machinerie de production est similaire. La biogénèse des miARNs et siARN chez divers organismes modèles pluricellulaires, le vertebré *Homo sapiens*, le vers *Caenorhabditis elegans*, l'insecte *Drosophila melanogaster* et la plante *Arabidopsis thaliana* montre de nombreuses similarités impliquant notamment une ou plusieurs endoribonuclease III de type Dicer associée à des cofacteurs protéiques. Quasiment toutes les protéines en jeu dans ces voies de biogenèses contiennent des domaines de liaison aux ARNs double brin (dsRBD). TRBP est l'un des cofacteurs de l'unique Dicer humaine et est composé de 3 dsRBDs. Bien que les dsRBDs soient composés de 3 feuillets β encadré de 2 hélices α il existe une variabilité dans leur structure et leurs fonctions cellulaires. Une comparaison des séquences et des structures des dsRBDs, à l'état libre, impliqué dans la biogenèse des miARNs and siARNs est réalisée. Les dsRBDs peuvent conférer de multiples fonctionnalités aux protéines qui les contiennent. Ils peuvent moduler l'activité des enzymes auxquels ils sont associés, peuvent avoir un effet sur la précision du clivage chez les plantes, influencent la production d'isoformes et leur distribution chez les metazoaires, peuvent avoir des rôles dans la sélection des substrats par les enzyme RNase III, semblent contribuer au moins faiblement à la sélection du brin chargé sur les ARNs messagers et certains dsRBD associés aux enzymes ADAR interagissent spécifiquement avec des séquences d'ARN double brin, contrairement à la majorité des dsRBDs. Les dsRBDs sont souvent présents en multiple copie, très souvent en tandems. Il y a souvent dans ce cas des différences d'affinité entre les dsRBDs dont l'origine est souvent obscure (zone d'interactions, dynamique). Le travail de thèse est focalisé sur le cofacteur TRBP.

20.1.3 Résumé de la partie II

La partie II explicite la stratégie et les principes et limites des méthodes biophysiques utilisés dans l'étude expérimentale. La construction de TRBP utilisée est celle correspondant à la région d'interaction avec les ARNs. Cette région correspond aux 228 premiers résidus de TRBP et comprends deux domaines de liaisons aux ARN doubles brins (dsRBD) séparée par une région interdomaine d'environ 60 résidus. Elle est notée TRBP-D1D2. Afin de pouvoir étudier le comportement des deux dsRBD dans la région de TRBP intéragissant avec les ARNs doubles brins, deux constructions réduites aux deux dsRBDs ont été produites (notées TRBP-D1 et TRBP-D2).

La comparaison du comportement de TRBP-D1D2 avec TRBP-D1 et TRBP-D2 permet d'obtenir des informations sur le comportement des dsRBDs dans TRBP-D1D2. Les précurseurs de microARN utilisés dans cette étude sont les deux précurseurs cytoplasmiques du microARN oncogène humain pre-miR-155 correspondant aux états avant et après clivage par l'enzyme Dicer. Afin que des conclusions fiables puissent être générées, le système a été analysé par plusieurs techniques biophysiques sélectionnées pour répondre aux questions posées tout en étant compatibles c'est à dire applicable sur le même système en solution et dans les mêmes conditions physico-chimiques. Pour chaque technique biophysique, les propriétés importantes pour comprendre l'origine du signal, les approximations principales utilisées pour démontrer les équations appliquées, ainsi que les protocoles utilisés dans cette thèse sont présentées. De plus, le développement s'attache à présenter différents phénomènes pouvant générer de façon non triviale des résultats indistinguables. La connaissance de ces processus est particulièrement utilisée lors de l'analyse des résultats expérimentaux. Les méthodes présentées incluent la chromatographie d'exclusion stérique, la diffusion de lumière, la calorimetrie isotherme, l'ultra-centrigugation analytique et plusieurs expériences de résonance magnétique nucléaire.

20.1.4 Résumé de la partie III

La partie III présente les méthodes de production des macromolecules biologiques utilisées dans le travail de thèse : les constructions de TRBP et des précurseurs de miR-155. La production des constructions de TRBP a été faite dans *E. coli* en milieu mimimum afin de faciliter les marquages pour la RMN. Pour augmenter les rendements de protéine en milieu deutéré, une sélection de colonie a été réalisée comme préalablement décrit (250). Une recette de lyse chimique des cellules a été mise au point pour augmenter la quantité de protéine récupérée à partir des cellules d'*E. coli*. Les protéines ont été purifiées par purification d'affinité sur colonne de Nickel avec des versions étiquetées des protéines comprenant une répétition d'histidines. Les étiquettes utilisées pour la purification d'affinité ont été systématiquement retirées pour éviter un possible biais. La dernière étape de purification est une chromatographie d'exclusion stérique. Les précurseurs de microARN 155 ont été synthetisés *in vitro* avec l'ARN polymérase du phage T7 avec par la plateforme cell-free de l'Institut de Biologie Structurale.

20.1.5 Résumé de la partie IV

La partie IV présente l'étude des constructions de TRBP et des précurseurs cytoplasmique de miR-155 à l'état isolé, ce qui est un préalable à leur étude en mélange (cf partie IV, complexe protéine/ARN). La combinaison de l'analyse des constructions protéiques par SDS-PAGE et par chromatographie d'exclusion stérique (SEC) a permis de valider leur homogénéité. Des expériences de SEC à plusieurs concentrations ont montré qu'il n'y a qu'un seul type d'assemblage en solution et leur analyse par diffusion de lumière à angles multiples (MALLS) a montré que ces constructions sont monomériques. La même conclusion est obtenue avec TRBP-D1D2 par ultracentrifugation analytique (svAUC), une méthode avec des approximations très différentes de celles du MALLS. L'analyse par svAUC montre aussi que TRBP-D1D2 a une forme allongée en solution. La structure secondaire des différentes constructions a ensuite été analysée par RMN. La structure secondaire et la flexibilité des dsRBDs isolés (TRBP-D1 et TRBP-D2) ou dans la construction double sont identiques. La région interdomaine de 60 résidus est entièrement flexible. Les deux dsRBDs 1 et 2 ont une structure majoritaire typique des dsRBDs et compatible avec les deux structures rapportées dans la littérature dans un cristal et en solution respectivement (290). Cependant le premier dsRBD subit un échange intégral plié-déplié (80%-20%) dans les constructions TRBP-D1 et TRBP-D1D2. Cet échange existe dans une large gamme de pH et de températures. L'analyse comparée des déplacements chimiques des dsRBDs isolés montre que les deux domaines n'intéragissent pas dans la construction TRBP-D1D2 : les deux domaines ont des conformations independantes. Une étude similaire réalisée sur le précurseur pre-miR-155 a montré que l'ARN produit est homogène et essentiellement monomérique bien qu'environ 4% des molécules soient présentes dans un état d'association plus gros, probablement un dimer. L'analyse par svAUC revèle que l'ARN a une forme allongée en solution. L'attribution des protons iminos par RMN a permis de confirmer que la conformation de l'ARN en solution est une tige-boucle typique des précurseurs de microARN. Ces informations sur les conformations en solution des constructions de TRBP et de pre-miR-155 ont servi de base pour l'étude de leurs intéractions.

20.1.6 Résumé de la partie IV

La partie IV présente l'étude de l'interaction entre la région de TRBP interagissant avec les ARNs (TRBP-D1D2) et les deux précurseurs cytoplasmiques du microARN miR-155. En premier lieu, l'utilisation combinée de la calorimetrie isotherme et de l'ultracentrifugation analytique a permis de construire un modèle macroscopique de cette interaction. Ce modèle indique que pour les deux précurseurs cytoplasmiques de miR-155, deux TRBP-D1D2 peuvent interagir successivement avec l'ARN avec des affinités similaires. La calorimetrie isotherme permet de mettre en évidence les deux associations successives, ce qui est confirmé par ultracentrifugation. Des modélisations du signal d'ultracentrifugation montrent qu'il est compatible avec les constantes de dissociation estimées par ITC. L'analyse de l'interaction par RMN montre que dans le complexe avec 1 TRBP-D1D2 interagissant avec l'un ou l'autre des précurseurs cytoplasmiques du microARN 155, les deux dsRBDs sont impliqués dans l'interaction et que la région interdomaine reste flexible et n'est pas impliquée dans l'interaction tout comme l'état déplié du premier dsRBD. Cette étude est la première analyse en détail de l'intéraction entre TRBP et des précurseurs de microARN. Les principales découvertes sont que la zone d'interaction avec les ARNs est restreinte aux dsRBDs, les deux dsRBDs sont impliqués dans la liaison avec une surface d’interaction canonique mais des dynamiques différentes, les deux premiers domaines de TRBP ont des conformations indépandentes dans les formes libres et liées avec les ARNs et deux TRBP-D1D2 peuvent lier une même molécule d’ARN. Les données suggèrent notamment que le rôle de TRBP peut être similaire avant et après le clivage du premier précurseur cytoplasmique par Dicer. Ces propriétés ouvrent la voie à une meilleure compréhension du rôle du cofacteur TRBP dans la biogenèse des microARN humains.

20.2 Assignement Note

Biomol NMR Assign
DOI 10.1007/s12104-012-9416-8

ARTICLE

Backbone resonance assignments of the micro-RNA precursor binding region of human TRBP

Matthieu P. M. H. Benoit · Michael J. Plevin

Received: 19 March 2012 / Accepted: 24 July 2012

Abstract TAR-RNA binding protein (TRBP) is a multi-domain human protein involved in micro-RNA (miRNA) biogenesis. TRBP is a component of both the Dicer complex, which processes precursor miRNAs, and the RNA-induced silencing complex-loading complex. In addition, TRBP is implicated in the human immunodeficiency virus replication cycle and interferon-protein kinase R activity. TRBP contains 3 double-stranded RNA binding domains the first two of which have been shown to interact with miRNA precursors. Here we present the backbone resonance assignments and secondary structure of residues 19–228 of human TRBP2.

Keywords TRBP · Triple resonance NMR experiments · Micro-RNA · Double-stranded RNA binding domain

Biological context

The majority of human micro-RNAs (miRNAs) are synthesized via a common two-step maturation pathway (Kim et al. 2009). Human miRNAs are transcribed as 1–2 kilobase (kb) primary transcripts. The 20–25 nucleotide (nt) miRNA sequence is located in a 60–70 base-pair (bp) stem-loop structure which is recognized and excised by the nuclear microprocessor complex. The resulting hairpin precursor miRNA (pre-miRNA) is exported to the cytoplasm where is it further processed by the Dicer complex. Cleavage of the pre-miRNA produces an imperfect double-stranded (ds) RNA duplex which is transferred from Dicer to Ago2 to form RISC, the protein complex responsible for miRNA-directed RNA interference.

Each miRNA maturation step is performed by a multi-protein complex that contains an RNase III enzyme and one or more accessory proteins. In humans the enzymatic component of the microprocessor, Drosha, interacts with a dsRNA binding domains (dsRBD)-containing protein, DGCR8, while Dicer has been shown to interact with two homologous proteins, HIV-1 transactivation response (TAR) element binding protein (TRBP) and/or PACT (Kim et al. 2009). In other organisms, the maturation of miRNAs or short interferring RNAs (siRNAs) is performed by protein complexes with similar compositions of homologous proteins.

TRBP was originally discovered via an interaction with HIV-1 TAR RNA and has subsequently shown to be implicated in the regulation of protein kinase R (PKR) activity. Two isoforms of TRBP exist, TRBP1 and TRBP2. The only difference between these isoforms is that TRBP2 has an additional N-terminal 21 residues. TRBP contains 3 dsRBDs and exhibits high sequence homology with PACT, which also regulates PKR. Both TRBP and PACT have been shown to associate with PKR via interactions mediated by dsRBDs, but the two proteins have opposite functional roles: TRBP inhibits PKR activity while PACT stimulates it.

Despite a clear involvement, the precise role of TRBP in miRNA biogenesis remains to be determined. TRBP has

---

M. P. M. H. Benoit · M. J. Plevin (✉)
Institut de Biologie Structurale Jean-Pierre Ebel, CEA, 41, rue Jules Horowitz, 38027 Grenoble, France
e-mail: michael.plevin@ibs.fr

M. P. M. H. Benoit · M. J. Plevin
Institut de Biologie Structurale Jean-Pierre Ebel, CNRS, 41, rue Jules Horowitz, 38027 Grenoble, France

M. P. M. H. Benoit · M. J. Plevin
Institut de Biologie Structurale Jean-Pierre Ebel, Université Joseph Fourier—Grenoble 1, 41, rue Jules Horowitz, 38027 Grenoble, France

been shown to interact with Dicer and components of the RISC complex. The first two dsRBDs of TRBP interact with miRNAs (Chakravarthy et al. 2010) and siRNAs (Yamashita et al. 2011). Four serine phosphorylation sites exist in TRBP all of which are located in the inter-domain linker regions (Chakravarthy et al. 2010). Phosphorylation of TRBP by the mitogen-activated protein kinase (MAPK) Erk has been shown to stabilise the interaction between TRBP and Dicer but not significantly affect the dicing activity of the complex.

Three dimensional (3D) structures of the first two dsRBDs of TRBP have been reported (PDB code: 2CPN, 3LLH). The 3D structure of dsRBD-2 in complex with a 10 bp siRNA duplex has also been elucidated (PDB code: 3ADL). These 3D structures reveal canonical dsRBD structures. Comparison of the 3D structure of dsRBD-2 alone and in complex with dsRNA reveals only minor RNA-induced conformational changes. To date only low resolution structural studies of larger constructs of TRBP have been reported and, despite several NMR studies of TRBP, no resonance assignments have been published. Here we report backbone ^{1}H, ^{13}C and ^{15}N chemical shifts and heteronuclear ^{1}H–^{15}N NOE values for a 210 residue fragments of human TRBP2 containing the two dsRBDs that interact with miRNA precursors, and 2 serine phosphorylation sites.

Methods and experiments

Expression and purification of [U-^{2}H, ^{13}C, ^{15}N]-labeled TRBP(19–228)

A codon-optimised cDNA encoding residues 19–228 of TRBP was cloned in to pLX06 plasmid (Protein'eXpert) to generate a construct with an N-terminal 6-histidine purification tag and a TEV cleavage site. The plasmid was transformed into *Escherichia coli* BL21(DE3) cells.

The production of deuterated TRBP was optimised using a double colony selection protocol in D_2O-based media as previously described (Sivashanmugam et al. 2009). The selected colony was used to inoculate pre-cultures of Luria Broth media (LB) prepared in D_2O and incubated at 37 °C with agitation (180 rpm) for 6 h. The LB pre-culture was used to inoculate 200 mL of [U-^{2}H, ^{13}C, ^{15}N]-labeled M9 minimal media to an initial optical density (OD) at 600 nm, OD_{600}, of 0.1. This second preculture was incubated overnight at 37 °C with agitation (180 rpm). M9 media was prepared with 1 g/L of [U-^{15}]-labeled NH_4Cl (Cambridge Isotope Laboratories) and 3 g/L of [U-^{2}H,^{13}C]-labeled glucose (Cambridge Isotope Laboratories) using [99.85 %-^{2}H]-labeled H_2O (Euriso-top). 2 L [^{13}C,^{15}N]-labeled M9 media prepared in D_2O was inoculated with sufficient pre-culture to give a starting $OD_{600} = 0.1$. The cultures were incubated at 37 °C with agitation (180 rpm) until OD_{600} reached 0.7 and then stored at 25 °C for 45 min. Protein expression was induced with IPTG to a final concentration of 1 mM. The induced cultures were incubated at 15 °C for 18 h. The final culture was harvested by centrifugation and the cell pellet stored at −80 °C until required.

TRBP(19–228) was purified at room temperature in protonated solvents to allow exchange of amide protons. Cell pellets were resuspended in lysis buffer (50 mM HEPES pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 10 mM imidazole, 10 % (v/v) glycerol, 1 mg/mL lysozyme, 0.5 % triton X-100, 1 μg/mL leupeptin, 1 μg/mL pepstatin, 20 μg/mL DNase and 20 μg/mL RNase). The volume, V, of lysis buffer required was determined by the final OD_{600} of the expression culture such that, $V(\text{buffer}) = OD_{600}(t_{max}) \times V(\text{culture})/40$. Cell lysis was performed at 30 °C for 40 min with agitation (180 rpm). 5 mM β-mercaptoethanol was then added and the lysis reaction clarified by centrifugation (30,000g, 30 min, 4 °C). The supernatant was passed over Ni-NTA affinity chromatography resin (2 mL of resin per L of bacterial culture; QiaGEN) pre-equilibrated with storage buffer (50 mM HEPES pH 7.5, 100 mM KCl, 10 mM $MgCl_2$, 10 mM $CaCl_2$, 10 mM imidazole, 10 % (v/v) glycerol and 5 mM β-mercaptoethanol). The resin was washed sequentially with 20 column volumes (CV) of storage buffer with 5 mM imidazole, 20 CV of storage buffer with 2 M NaCl and 20 CV of storage buffer alone. Bound protein was eluted with 2.5 CV of storage buffer with 250 mM imidazole. The eluted protein was incubated with 0.1 equivalents of TEV protease (stored in 50 mM Tris–HCl pH 8.0, 25 mM NaCl, 0.5 mM EDTA, 2 mM DDT and 20 % glycerol). The reaction mixture was dialysed against 2 × 2 L of storage buffer at 4 °C (12 h then 4 h). The dialysate was passed over 1 CV of pre-equilibrated (storage buffer) Ni-NTA affinity chromatography resin and then washed with storage buffer. Fractions containing protein were concentrated and 500 μL aliquots were passed over an analytical Superdex S75 10/300 GL column (GE Healthcare) pre-equilibrated with gel filtration buffer (23.5 mM potassium phosphate, pH 6.5, 100 mM KCl, 10 mM $MgCl_2$ and 5 mM β-mercaptoethanol). The chromatograms were characterized by a single well resolved peak (data not shown). The protein-containing fractions were pooled and concentrated. [^{15}N]-labeled samples were produced with the same protocol using H_2O and unlabeled glucose. NMR data of TRBP(19–228) were collected in the presence of 5 mM β-mercaptoethanol or 2–5 mM tris(2-carboxyethyl)phosphine (TCEP)

NMR spectroscopy

Three dimensional (3D) NMR experiments for backbone assignment were recorded using 90 μL of a 600 μM

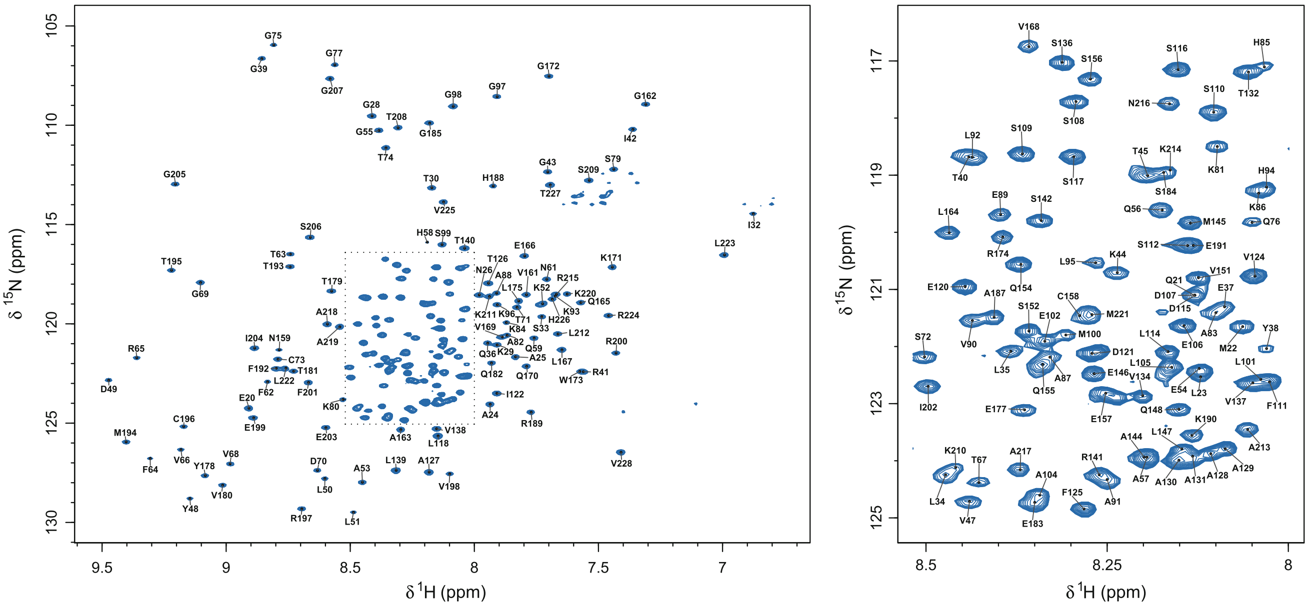


Fig. 1 (^{1}H,^{15}N) TROSY-HSQC spectrum of [^{15}N]-labeled TRBP (19–228). ^{15}N–^{1}H^{N} resonance assignments are indicated. The *panel* on the *right* shows an expansion of the region of the full spectrum contained in the *dashed box*. The spectrum was acquired on a Varian (Agilent) DirectDrive 800 MHz spectrometer at 25 °C in 23.5 mM phosphate buffer at pH 6.5 with 2 mM TCEP

sample of TRBP(19–228) in a 2.5 mm diameter Shegemi tube (Shigemi Inc.) placed in a 5 mm diameter standard tube filled with D$_2$O. 3D BEST TROSY versions of HNCACB, HNCOCACB and HNCO experiments (Favier and Brutscher 2011) were performed at 25 °C using a Varian (Agilent) DirectDrive 600 MHz spectrometer equipped with a triple resonance cryogenic probe. Heteronuclear ^{1}H–^{15}N NOE experiments (Farrow et al. 1994) were measured on a Varian (Agilent) DirectDrive 600 MHz at 25 °C with a room temperature probe using 300 μL of a 1.4 mM sample of [^{15}N]-labeled TRBP(19–228) in a 5 mm Shegemi tube (Shigemi Inc.) with 5 % D$_2$O. All NMR data were processed with NMRPipe (Delaglio et al. 1995) and analysed using CCPNMR v 2.1.5 (Vranken et al. 2005). Heteronuclear NOE values of well-resolved crosspeaks were calculated as previously described (Farrow et al. 1994).

Resonance assignment and data deposition

It was possible to obtain assignments for 209 out of 210 residues of TRBP(19–228), including 99.5 % of non proline ^{15}N–H^{N} pairs and ^{13}C$^{\alpha}$, ^{13}C$^{\beta}$ chemical shifts and 91 % of ^{13}C′ chemical shifts. An example of a (^{1}H,^{15}N) TROSY-HSQC spectrum of [^{13}C,^{15}N]-labeled TRBP(19–228) with sequence-specific assignments is given in Fig. 1.

Resonance assignments of TRBP(19–228) have been deposited in the BioMagResBank (BMRB; http://www.bmrb.wisc.edu) under the BMRB accession number 18324.

Secondary structure and backbone flexibility

Secondary structure predictions of TRBP(19–228) were calculated using TALOS+ (Shen et al. 2009) and the results compared with the 3D structures of dsRBD-1 (PDB

Table 1 NMR experiments recorded of TRBP(19–228)

Sample	Experiment	Recycle delay [s]	Number of scans	t^{max}(H) [ms]	t^{max}(N) [ms]	t^{max}(carbon)
Protonated	BEST-TROSY-HSQC	0.5	120	70	80	na
Deuterated	BEST-TROSY-HSQC	1.5	8	70	75	na
Deuterated	BEST-TROSY-HN(CO)CACB	1.5	8	64	26	8.3 (CACB)
Deuterated	BEST-TROSY-HNCACB	1.5	8	64	24	8.3 (CACB)
Deuterated	BEST-TROSY-HNCO	1.5	4	64	26	10 (CO)
Protonated	^{1}H–^{15}N-het-NOE	5	24	70	81	na

t^{max} maximum acquisition time

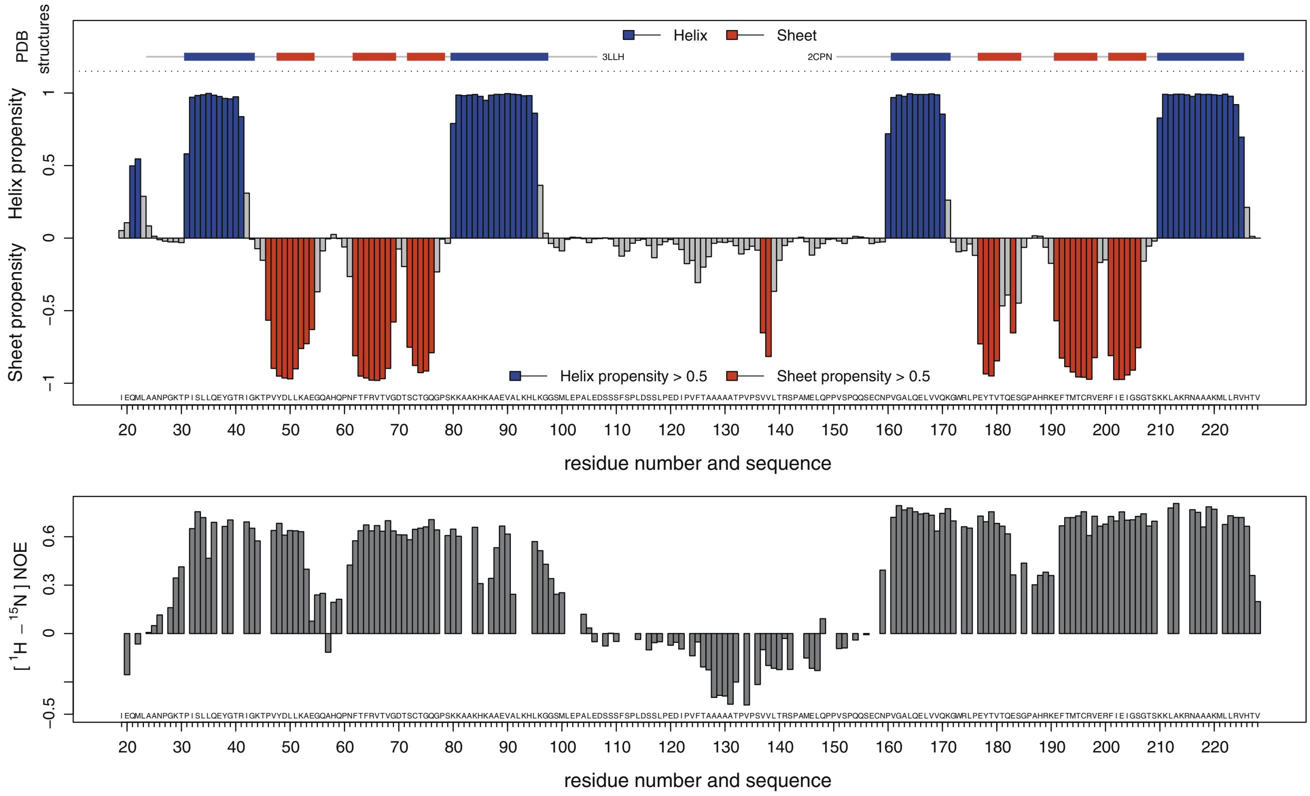


Fig. 2 *Top*: helical and sheet secondary structure propensity predictions of TRBP(19–228) calculated using TALOS+ compared to the secondary structure from previously reported 3D structures of dsRBD-1 (3LLH) and dsRBD-2 (2CPN). *Bottom*: heteronuclear ^{1}H–^{15}N NOE values of TRBP(19–228)

code: 3LLH) or dsRBD-2 (2CPN) in Fig. 2. The secondary structure prediction obtained with data collected from a construct containing both dsRBDs is in excellent agreement with the previously reported 3D structures of the two individual domains (Yamashita et al. 2011).

In order to characterize the flexibility of TRBP(19–228), we measured heteronuclear ^{1}H–^{15}N NOE values, which provide information about the motions of ^{1}H–^{15}N bond vectors on the ns-ps time-scale. The high ^{1}H–^{15}N NOE values in the secondary structure elements of both domains confirm that the two dsRBDs are folded. The linker connecting the two dsRBDs is characterized by negative heteronuclear NOE values indicating a high degree of flexibility. The only positive heteronuclear NOE value in the linker region concerns a residue preceding two adjacent prolines, a motif that is known to add local order to polypeptide chains.

Finally, each domain contains a short stretch of decreased heteronuclear NOE values that centers on the conserved histidine of the loop between β-strands 1 and 2 (residues 53–61, dsRBD-1, and residues 183–191, dsRBD-2). This loop comprises the second of three regions in dsRBDs that form direct or water-mediated contacts with dsRNAs (Tian et al. 2004). The enhanced flexibility of this loop compared to the rest of the domain may indicate a role in target recognition.

Acknowledgments The authors would like to thank Drs. A. Favier and E. Condamine for assistance with NMR data acquistion, Drs. B. Bersch and J. Boisbouvier for helpful commments, and the Partnership for Structural Biology, Grenoble, for access to integrated structural biology platforms. This work was supported in part by funding from the E. U. (FP7-PEOPLE-IRG-2008) and l'Agence pour la Recherche sur le Cancer to MJP. MPMHB acknowledges the CEA for a Ph. D. studentship.

References

Chakravarthy S, Sternberg SH, Kellenberger CA, Doudna JA (2010) Substrate-specific kinetics of Dicer-catalyzed RNA processing. J Mol Biol 404:392–402

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293

Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. Biochemistry 33:5984–6003

Favier A, Brutscher B (2011) Recovering lost magnetization: polarization enhancement in biomolecular NMR. J Biomol NMR 49:9–15

Kim VN, Han J, Siomi M (2009) Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol 10:126–139

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44:213–223

Sivashanmugam A, Murray V, Cui C, Zhang Y, Wang J, Li Q (2009) Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. Protein Sci 18:936–948

Tian B, Bevilacqua PC, Diegelman-Parente A, Mathews MB (2004) The double-stranded-RNA-binding motif: interference and much more. Nat Rev Mol Cell Biol 5:1013–1023

Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59:687–696

Yamashita S, Nagata T, Kawazoe M, Takemoto C, Kigawa T, Guntert P, Kobayashi N, Terada T, Shirouzu M, Wakiyama M, Muto Y, Yokoyama S (2011) Structures of the first and second double-stranded RNA-binding domains of human TAR RNA-binding protein. Protein Sci 20:118–130

20.3 Article on TRBP-D1D2 interactions with miR-155 precursors

Nucleic Acids Research Advance Access published February 21, 2013

Nucleic Acids Research, 2013, **1–12**
doi:10.1093/nar/gkt086

The RNA-binding region of human TRBP interacts with microRNA precursors through two independent domains

Matthieu P. M. H. Benoit[1,2,3], Lionel Imbert[1,2,3], Andrés Palencia[4,5,6], Julien Pérard[4,5,6], Christine Ebel[1,2,3], Jérôme Boisbouvier[1,2,3] and Michael J. Plevin[1,2,3,*]

[1]CEA, Institut de Biologie Structurale Jean-Pierre Ebel, Grenoble, France, [2]CNRS, Institut de Biologie Structurale Jean-Pierre Ebel, Grenoble, France, [3]Université Joseph Fourier – Grenoble 1, Institut de Biologie Structurale Jean-Pierre Ebel, Grenoble, France, [4]European Molecular Biology Laboratory, Grenoble Outstation, Grenoble, France, [5]Université Joseph Fourier – Grenoble 1, Unit of Virus Host-Cell Interactions, Grenoble, France and [6]CNRS, Unit of Virus Host-Cell Interactions, Grenoble, France

Received April 13, 2012; Revised January 21, 2013; Accepted January 23, 2013

ABSTRACT

MicroRNAs (miRNAs) are small non-coding RNAs that regulate gene expression through RNA interference. Human miRNAs are generated through a series of enzymatic processing steps. The precursor miRNA (pre-miRNA) is recognized and cleaved by a complex containing Dicer and several non-catalytic accessory proteins. HIV TAR element binding protein (TRBP) is a constituent of the Dicer complex, which augments complex stability and potentially functions in substrate recognition and product transfer to the RNA-induced silencing complex. Here we have analysed the interaction between the RNA-binding region of TRBP and an oncogenic human miRNA, miR-155, at different stages in the biogenesis pathway. We show that the region of TRBP that binds immature miRNAs comprises two independent double-stranded RNA-binding domains connected by a 60-residue flexible linker. No evidence of contact between the two double-stranded RNA-binding domains was observed either in the apo- or RNA-bound state. We establish that the RNA-binding region of TRBP interacts with both pre-miR-155 and the miR-155/miR-155* duplex through the same binding surfaces and with similar affinities, and that two protein molecules can simultaneously interact with each immature miRNA. These data suggest that TRBP could play a role before and after processing of pre-miRNAs by Dicer.

INTRODUCTION

The majority of human micro-RNAs (miRNAs) are produced from long RNA transcripts in a two-step process (1,2). The initial primary miRNA (pri-miRNA) transcript can be 1-2 kilobases (kb) in length. Within the pri-miRNA, a short, imperfect stem-loop structure is recognized and excised by the nuclear 'Microprocessor' complex. The product is a 60–70-nucleotide (nt) precursor miRNA (pre-miRNA), which is exported from the nucleus by an Exportin-5/Ran-GTP complex. In the cytoplasm, the pre-miRNA is further processed by the Dicer complex into a 20–25-base-pair (bp) duplex, which contains the final miRNA and its reverse complement, termed miRNA*. The imperfect miRNA/miRNA* duplex is then transferred to the RNA-induced silencing complex (RISC), which is the mediator of miRNA-directed RNA interference in humans.

The multidomain human TAR element binding protein (TRBP) is a component of the Dicer complex and has been implicated in the processing of pre-miRNAs (3,4). TRBP was initially discovered as a protein that associates with the HIV TAR RNA (5) and was later demonstrated to interact with and regulate interferon-induced protein kinase R (PKR), a kinase involved in the cellular response to viral infection (6,7). In the context of miRNA biogenesis, TRBP associates with Dicer, the RNase III enzyme that catalyses the removal of the apical loop of pre-miRNAs (8) and processes small interfering RNAs [siRNAs; (9)]. TRBP has been shown to enhance dicing and augment Dicer stability (3,4,10), although the molecular mechanisms for these activities have not been elucidated. Downstream of Dicer, TRBP

---

*To whom correspondence should be addressed. Tel: +44 1904 328682; Email: michael.plevin@york.ac.uk
Present address:
Michael J. Plevin, Department of Biology, University of York, York YO10 5DD, UK.

is also a component of the RISC-loading complex (RLC), the protein assembly responsible for transferring the miRNA/miRNA* duplex from Dicer to RISC (11,12). TRBP contains three double-stranded (ds) RNA-binding domains (dsRBDs), but only the first two dsRBDs interact with dsRNA. The third dsRBD is believed to mediate protein/protein interactions and bind to other proteins in the Dicer complex. The contribution of TBRP to the function of either the Dicer complex or RLC is not well-understood. The presence of two types of dsRBD supports the hypothesis that TRBP mediates both protein/RNA and protein/protein contacts, although the targets, location and consequences of these interactions in the context of the Dicer complex or RLC remain to be elucidated.

Dicer homologues are found in *Drosophila melanogaster* and *Caenorhabditis elegans*, and they associate with proteins with a similar domain composition to TRBP. Proteins such as *R2D2* (13) and *Loquacious* (14,15) in *D. melanogaster* and RDE-4 (16) in *C. elegans* all contain two N-terminal canonical RNA-binding dsRBDs as well as a non-canonical C-terminal dsRBD, which is implicated in protein–protein interactions with the respective Dicer homologue. In plants, Dicer-like 1 (DCL1) is found in complex with HYPONASTIC LEAVES 1 (HYL1), an accessory protein, which contains only two dsRBDs (17). HYL1 interacts with pre-miRNA and miRNA/miRNA* duplexes primarily through its first dsRBD (18,19). Human Dicer has also been shown to associate with a TRBP homologue, Protein Activator of PKR [PACT; (20,21)]. A complex containing Dicer, TRBP and PACT can be purified from human cells (21). PACT also contains three dsRBDs and, like TRBP, is implicated in the regulation of PKR. However, unlike TRBP, PACT activates PKR (22).

miRNAs regulate gene expression at the post-transcriptional level (1,2). To date, ~1500 human miRNAs have been documented in the database of miRNAs, miRbase (23). miRNAs are believed to regulate the expression of upwards of 30% of human genes by enabling RISC to recognize complementary sequences in target messenger RNAs. Many miRNAs have oncogene- or tumor-suppressor-like behaviour (24). Mis-expression or mutation of miRNAs can have considerable cellular consequences (25). For example, miR-155 has been implicated in lung and breast cancer as well as different leukaemias and lymphomas (24). Mutations in the proteins involved in the processing or regulation of miRNA biogenesis have also been linked to different cancers. For example, certain frame-shift mutations implicated in colorectal or endometrial cancers result in truncated forms of TRBP that lack dsRBDs 2 and 3 (26). Furthermore, TRBP is phosphorylated by the mitogen-activated protein kinase Erk (27), and it has been suggested to interact with the anti-cancer drug enoxacin (28). These recent studies have provided further evidence of the importance of TBRP in miRNA biogenesis.

Over the past decade, there has been a drive to understand the structural biology of miRNA biogenesis (29,30). Nuclear magnetic resonance (NMR) spectroscopic and X-ray crystallographic studies of TRBP have elucidated the isolated 3D structures of the first two dsRBDs (31) as well as the second dsRBD in complex with a short siRNA duplex (18). Negative-stain electron microscopy studies of miRNA biosynthesis complexes containing TRBP have also been reported; however, the level of resolution is not yet sufficient to accurately describe the location of TRBP in these complexes or determine how it interacts with other components (32,33). Many biochemical studies have sought to explore the mechanism for RNA recognition by TRBP or examine how TRBP associates with Dicer (3,4,10–12,34,21). However, despite these efforts, the exact role of TRBP in immature miRNA recognition and processing remains unclear.

Here we present an in-depth biophysical analysis of the complex formed between the RNA-binding region of TRBP and biosynthetic precursors of the oncogenic human miRNA miR-155. We establish that the two dsRBDs that form the RNA-binding region function independently in apo- and RNA-bound forms. We show that this region of TRBP interacts with both pre-miR-155 and miR-155/miR-155* with affinities in the low micromolar range and that it can form a large complex in which four dsRBDs interact with a single immature miRNA molecule. These data, which concern functional regions of the protein and their interactions with relevant biological targets, further enhance our understanding of the role of TRBP in miRNA biogenesis.

MATERIALS AND METHODS

Protein expression and purification

Residues 19–99 (TRBP-D1; monomeric molecular weight: 8.7 kDa), 157–228 (TRBP-D2; 8.2 kDa) or 19–228 (TRBP-D12; 22.6 kDa) of TRBP were sub-cloned into pLX06 (Protein'eXpert, France) using a codon-optimized TRBP2 (AAA36765) cDNA template (GeneArt). Each construct was expressed with an N-terminal 6-histidine purification tag and a tobacco etch virus cleavage site. Protein constructs were expressed and purified as previously described (35).

Large-scale production of pre-miR-155 and miR-155/miR-155*

Pre-miR-155 (5'-GGAAU GCUAA UCGUG AUAGG GGUUU UUGCC UCCAA CUGAC UCCUA CA UAU UAGCA UUCCC A-3') or miR-155/miR-155* (miR-155: 5'-GGAAU GCUAA UCGUG AUAGG CCU-3'; miR-155*: 5'-GGCCU ACAUA UUAGC AUUCC CA-3') were produced by *in vitro* transcription using in-house produced T7 RNA polymerase and fully complementary ds DNA oligonucleotides containing a T7 promoter. As each translated RNA sequence was required to start with a GG dinucleotide, it was necessary to modify the native RNA sequences. Additional sequence changes were required to preserve the Watson–Crick base-pairing. Non-native sequence nucleotides are underlined in the sequences provided previously.

Template oligonucleotides (Eurogentec) were annealed by incubation at 95°C for 5 minutes. The transcription reaction was performed in a standard T7 RNA polymerase buffer (40 mM Tris-HCl pH 8.1, 1 mM spermidine, 5 mM dithiothreitol, 0.1 % Triton X-100, 40 µg/ml bovine serum albumin, 80 mg/ml polyethylene 8000, 4 mM of each nucleotide triphosphate, 15-20 mM $MgCl_2$, 35 µg/ml DNA matrix, 0.1 mg/ml T7 RNA polymerase and 1 U/ml pyrophosphatase). The reaction mixture was incubated at 37°C for 3 hours. RNA products were purified by polyacrylamide gel electrophoresis, dialyzed multiple times against water and lyophilized. The folding of each RNA was achieved by heating each sample in water at 96°C for 5 minutes, followed by rapid cooling induced by the addition of ice-cold 2× NMR buffer. All RNA samples were analysed by 1D (^{1}H) NMR spectroscopy.

NMR spectroscopy

(^{1}H,^{13}C,^{15}N) backbone resonance assignments of TRBP constructs in complex with either pre-miR-155 or miR-155/miR-155* were obtained using a combination of 3D BEST triple resonance NMR experiments (36) and titration experiments. (^{1}H,^{13}C,^{15}N) backbone resonance assignments of apo-TRBP-D12 were taken from our previous work [BioMagResBank code: 18324; (35)]. Assignments of TRBP-D12 in complex with pre-miR-155 were obtained by analysing 3D (^{1}H,^{13}C,^{15}N) HNCA and HNCO experiments, which were acquired using a 230 µM [U-^{2}H,^{13}C,^{15}N]-labelled sample of TRBP-D12 and a 3.5-fold excess of unlabelled pre-miR-155. The results of titration experiments (see later in the text) were also used to aid assignment of TRBP-D12 in complex with pre-miR-155 or in complex with miR-155/miR-155*.

Apo- and RNA-bound states of the different TRBP constructs were analysed using 2D BEST (^{1}H,^{15}N) TROSY experiments at various protein concentrations (0.4 to 0.8 mM) and RNA/protein ratios. Titration experiments were performed with the following RNA/protein ratios: (i) TRBP-D12 with 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0 and 2.3 molar equivalents of pre-miR-155 or miR-155/miR-155*; (ii) TRBP-D1 with 0, 0.1, 0.22, 0.29, 0.33, 0.4, 0.5, 0.75, 1.0 and 1.5 molar equivalents of pre-miR-155; or (iii) TRBP-D2 with 0, 0.1, 0.5, 0.75, 1.0 and 1.5 molar equivalents of pre-miR-155. The time required to collect individual titration points depended on the protein concentration and the signal-to-noise ratio.

All NMR experiments were performed at 25°C using a Varian (Agilent) DirectDrive 800 MHz spectrometer equipped with a triple resonance cryogenic probe. NMR spectra were processed using NMRpipe (37) and analysed using CCPNMR Analysis, version 2.1.5 (38). Compound chemical shift differences, Δδ, between free protein and RNA-bound protein were calculated using the following equation: $\Delta\delta = [\Delta\delta_{\mathrm{H}}^2 + (\Delta\delta_{\mathrm{N}}/6.5)^2]^{1/2}$, where $\Delta\delta_{\mathrm{X}}$ is the different in chemical shift between comparable resonances in different 2D spectra (39).

Multi-angle laser light scattering

Fifty-microlitre samples of TRBP-D1 (1.8 mg/ml), TRBP-D2 (1.7 mg/ml), TRBP-D12 (2 mg/ml) and pre-miR-155 (4 mg/ml) were loaded onto an analytical Superdex S200 size exclusion chromatography (SEC) column (GE Healthcare) pre-equilibrated with NMR buffer. SEC was performed at 0.5 ml/min with an in-line multi-angle laser light scattering (MALLS) spectrometer (DAWN EOS, Wyatt Instruments) with 18 angles measuring at 690 nm. An in-line refractive index detector (RI 2000, Spares) was used to follow the differential refractive index relative to the solvent (1.34). Masses were estimated with the Debye model using ASTRA software version 5.3.4.20 (Wyatt Instruments) using a theoretical dn/dc value of 0.185 ml/g.

Sedimentation velocity analytical ultracentrifugation

Sedimentation velocity experiments were performed with a Beckmann Coulter XLI analytical ultracentrifuge at 50 000 rpm (185 000g), 25°C with an An-50Ti rotor. Samples or reference (NMR buffer) were loaded into sapphire-windowed cells with 12-mm optical paths, which had previously been treated with RNase ZAP (Sigma-Aldrich). Acquisition was performed at 260 nm and 280 nm, and using interference optics. The molar masses for the polymers were calculated from their composition. A partial specific volume ($\bar{v}$) of 0.74 ml/g was estimated for TRBP-D12 using the software SEDNTERP version 2 (40). The $\bar{v}$ used for pre-miR-155 (0.508 ml/g) was taken from a study of a viral RNA in dilute aqueous KCl solutions (41). The data were analysed in SEDFIT [Version 12.52; (42)] using the c(s) analysis and non-interacting species models. The latter was used to estimate buoyant molar masses, M_{b}, using the equation, $M_{\mathrm{b}} = M \times [1 - (\rho \times \bar{v})]$, where M is the molar mass and ρ the solvent density. Hybrid models and multi-wavelength analysis were performed in the program SEDPHAT (43). Results in terms of buoyant mass and multi-wavelength analysis are given for the more concentrated sample, which yielded higher signal-to-noise ratios. The mass density (1.005 g/ml) and viscosity (0.898 cP) of the buffer were calculated using SEDNTERP (v20120328 beta, www.jphilo.mailway.com).

The $\bar{v}$ for possible protein/RNA complexes was estimated by summing the individual $\bar{v}$ values of each component weighted by their fractional mass in the complex (Supplementary Table S1). The sedimentation coefficient, s, of each complex estimated from the distribution of sedimentation coefficients, c(s), was not affected by the choice of the three possible values. c(s) profiles reported here were fitted with a value of 0.64 ml/g. The Svedberg and Stokes equations were used to analyse the value of the experimental sedimentation coefficients.

Isothermal titration calorimetry

All isothermal titration calorimetry (ITC) experiments were conducted using a MicroCal ITC-200 at 25°C. ITC experiments were performed in two ways: either with the 'ligand', in this case the protein, in the cell, or with 'ligand' in the syringe. Protein concentrations varied between 30 and 60 µM in the sample cell and 250 and 500 µM in the syringe. RNA concentrations varied between 25 and 40 µM in the sample cell and 150 and 300 µM in the syringe. The heat evolved after each injection of protein or RNA was obtained from the integral of the baseline-corrected calorimetric signal. The equations

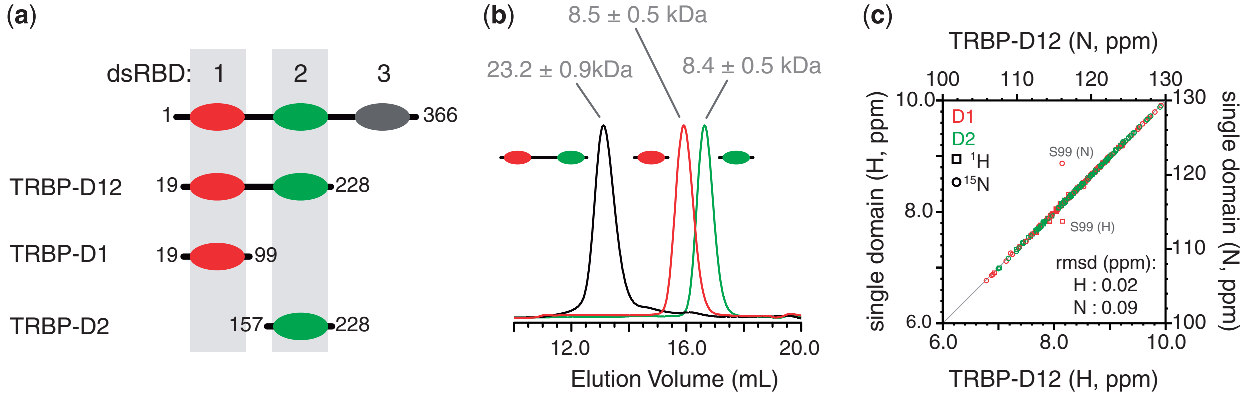


Figure 1. Analysis of the solution behaviour of dsRBDs from TRBP. (**a**) Diagram of the three constructs of TRBP used in this study. The same colour scheme for TRBP-D1 (red) and TRBP-D2 (green) is used throughout; (**b**) analytical SEC profiles showing elution volume *vs.* absorbance at 280 nm, A_{280}, for TRBP-D1 (red), TRBP-D2 (green) and TRBP-D12 (black). The molecular weight of each species determined by MALLS is given; (**c**) site-by-site comparison of the (^{1}H,^{15}N) chemical shift of a residue in the single domain and the corresponding residue in TRBP-D12. Circles denote (^{15}N) chemical shifts and squares (^{1}H) chemical shifts. Amide groups from TRBP-D1 are coloured in red and those from TRBP-D2 in green. The points corresponding to the large outliers are annotated in grey. RMSD values for ^{1}H and ^{15}N chemical shifts are provided.

used to analyse ITC data of TRBP-D12 were derived from the classical sequential binding equation using activity coefficients of 1. Previous assumptions of independency or co-operativity were not made. The resulting binding isotherms were fitted to maximize the likelihood of the data assuming a Gaussian noise distribution. The titration data were globally fitted for each pair of experiments using previously described protocols (44). The standard deviation of the estimation of each binding parameter was determined by computing 100 Monte Carlo simulations of each experiment performed assuming a Gaussian noise of 70 nanocals (estimated from our data) followed by a global fitting of the data. The stoichiometry for each protein/RNA pair was estimated by comparing the fits of sequential models for 1:1 and 1:2 protein/RNA complexes (see Supplementary Material).

RESULTS

Two independent dsRBDs comprise the RNA-binding region of TRBP

Three constructs derived from the RNA-binding region of TRBP were produced (Figure 1a) and characterized by SEC. Analysis of the eluted protein by SEC-MALLS reported values of 8.5 ± 0.5 kDa, 8.4 ± 0.5 kDa and 23.2 ± 0.9 kDa for TRBP-D1, TRBP-D2 and TRBP-D12, respectively (Figure 1b). These values correspond well with the respective monomeric masses of each protein construct.

Two dimensional (^{1}H,^{15}N) HSQC spectra of each of the three constructs displayed good ^{1}H resonance dispersion, which indicated that the dsRBDs adopt folded 3D structures. The spectrum of the double-domain construct, TRBP-D12, can be reconstituted by overlaying the spectra of TRBP-D1 and TRBP-D2 (Supplementary Figure S1a–d). The additional peaks visible in the spectrum of TRBP-D12 correspond to residues 96–151, which are located in the linker region. The (^{1}H,^{15}N) resonance frequencies of a given residue in a single-domain construct compare well with their counterparts in the double-domain construct, with root mean squared deviation (RMSD) values of 0.02 and 0.09 ppm for ^{1}H and ^{15}N frequencies, respectively (Figure 1c). The largest deviations in compound chemical shift differences between comparable sites are predictably found at the C-terminus of TRBP-D1 (Supplementary Figure S1e). These differences can be easily attributed to the presence of the inter-domain linker in the double-domain construct. The similarity in corresponding NMR signals strongly suggests that the two domains do not interact in the tandem dsRBD construct. This observation is consistent with previously reported (^{1}H,^{15}N) heteronuclear NOE data, which demonstrated the flexibility of the inter-domain linker (35). Taken together, these data suggest that the RNA-binding region of TRBP is composed of two independent dsRBDs connected by a flexible linker and, therefore, that it is similar to the tandem dsRBDs that form the dsRNA-interacting region of PKR (45).

Both dsRBDs of the RNA-binding region of TRBP interact with pre-miR-155

Large, site-specific changes in (^{1}H,^{15}N) chemical shifts are observed when unlabelled pre-miR-155 is added to constructs of TRBP (Figure 2a). Backbone resonance assignments were obtained of TRBP-D12 in the presence of an excess of pre-miR-155. Using these sequence-specific assignments, it was possible to evaluate which regions of each dsRBD were involved in the formation of protein/RNA complexes. A comparison of the amide chemical shifts of apo- and RNA-bound TRBP-D12 revealed large RMSD values between the two data sets (0.11 ppm for ^{1}H and 0.54 ppm for ^{15}N frequencies; Supplementary Figure S2). The largest chemical shift changes observed on complex formation are restricted to amino acids located in the two dsRBDs. Only negligible chemical shift changes occur in the linker region (Figure 2a–c). The profiles of residue-by-residue chemical shift changes of the two dsRBDs in TRBP-D12 are highly comparable (Figure 2a–c).

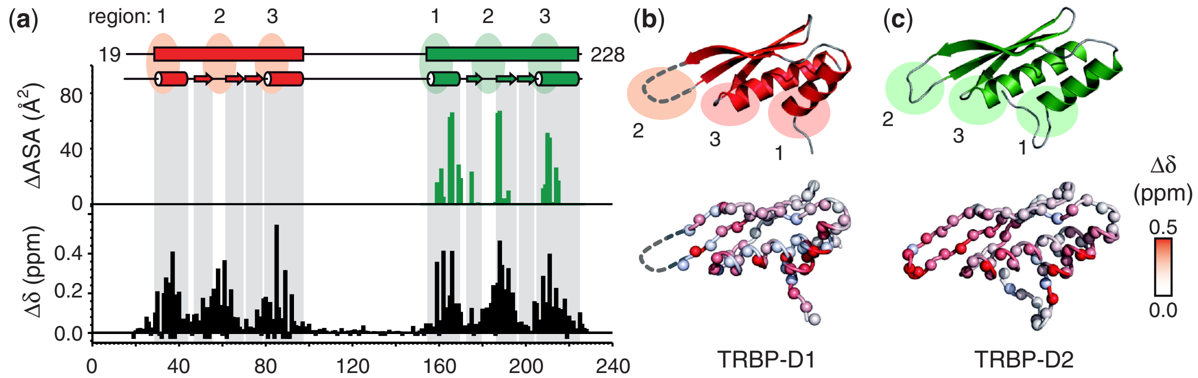


Figure 2. NMR analysis of the interaction between TRBP-D12 and pre-miR-155. (**a**) Compound chemical shift values (Δδ) measured when TRBP-D12 interacts with pre-miR-155 plotted as a function of residue number (black). $\Delta\delta = [\Delta\delta_H^2 + (\Delta\delta_N/6.5)^2]^{1/2}$, where $\Delta\delta_X$ is the difference in chemical shift between apo- and RNA-bound spectra (39). Prolines and unassigned residues have been given values of −0.05 and −0.1 ppm, respectively. For comparison, the change in accessible surface area (ΔASA) when TRBP-D2 forms a complex with dsRNA is shown (green). ΔASA were calculated using the POPS* server (46) from 3ADL (18). The boundaries of secondary structure elements were taken from the 3D structures of TRBP-D1 [3LLH; (31)] and TRBP-D2 [3ADL; (18)]; (**b,c**) TOP: cartoon representation of the 3D structures of TRBP-D1 and TRBP-D2 showing the three regions in canonical dsRBDs that are implicated in dsRNA binding; (b,c) BOTTOM: Δδ values plotted on the respective 3D structures. Each amide nitrogen is represented by a sphere and coloured according to the Δδ scale provided. The $\beta_{2/3}$ loop was not resolved in the 3D structure of TRBP-D1 and is shown by a broken line.

Three critical regions in canonical dsRBDs mediate the interaction with dsRNA (47): region 1, α-helix 1 (α_1); region 2, the loop between strands β_2 and β_3 ($\beta_{2/3}$); and region 3, the N-terminus of α_2 (Figure 2b and c). The NMR chemical shift perturbation profiles showed that both TRBP-D1 and TRBP-D2 interact with pre-miR-155 through the classical dsRBD interaction site. Residues located in regions 1, 2 and 3 exhibit the largest changes in chemical shift or disappear entirely from the spectrum, even after the addition of 2.3-fold excess of pre-miR-155 (Figure 2). These chemical shift changes are consistent with previously published structural and biochemical analyses of other dsRBD/dsRNA complexes (31, 47–50). The 3D structure of TRBP-D2 in complex with a short siRNA was reported in 2010 (18). The residue-by-residue changes in solvent accessible surface area, ΔASA, that occur on the formation of this complex highlight the location of the three critical RNA interaction regions in TRBP-D2 (Figure 2a). The profile of ΔASA for the TRBP-D2/siRNA complex and Δδ observed for dsRBD-2 in the TRBP-D12/pre-miR-155 complex correspond well. These observations suggest that the interaction between TRBP-D2 and the imperfect stem-loop of pre-miR-155 in solution is similar to that reported for a perfect A-form dsRNA duplex and, furthermore, that highly similar protein/RNA contacts are formed in the context of the tandem dsRBD construct.

Titration of [^{15}N]-labelled TRBP-D1 or TRBP-D2 with unlabelled pre-miR-155 revealed that both of the isolated dsRBDs of TRBP can interact with pre-miR-155. Two-dimensional (^{1}H,^{15}N) HSQC spectra of either TRBP-D1 or TRBP-D2 showed clear and significant chemical shift perturbations after the addition of RNA. NMR data from sub-equimolar titration points were characterized by severe line-broadening. Many cross-peaks broaden beyond detection after the addition of 0.1 equivalents of RNA and reappear when the RNA/protein ratio exceeds 0.3–0.5 (Supplementary Figure S3). Two-dimensional (^{1}H,^{15}N) HSQC spectra of TRBP-D12 in complex with pre-miR-155 superpose nicely with spectra of either of the single dsRBDs in complex with the same RNA (Figure 3a–d). As was the case for the apo-protein (Supplementary Figure S1), the spectrum of the double-domain construct in complex with pre-miR-155 is essentially a superposition of the spectra of the individual dsRBDs in complex with the same RNA (Figure 3a–d). Site-by-site differences in (^{1}H,^{15}N) resonance frequencies between single-domain and double-domain data sets are very small, with RMSD values of 0.02 and 0.11 ppm for ^{1}H and ^{15}N frequencies, respectively (Figure 3e). That is, a cross-peak for an amide group in the TRBP-D1/pre-miR-155 complex, for example, has almost exactly the same (^{1}H,^{15}N) resonance frequencies, and therefore experiences a near-identical local chemical environment, as the comparable amide group in the tandem domain/pre-miR-155 complex. This similarity of resonance frequencies between single- and double-domain constructs suggests that each dsRBD forms essentially the same interaction regardless of whether it is a component of the tandem domain construct. If the interaction with pre-miR-155 led to (or resulted from) the formation of contacts between the two dsRBDs, this would result in chemical shift differences between NMR spectra of the single- and double-domain constructs. However, the small RMSD between these data sets suggests that each dsRBD in TRBP-D12 interacts independently with pre-miR-155.

The RNA-binding region of TRBP can form a 2:1 complex with pre-miR-155

ITC experiments were conducted to further characterize the complex formed between the dsRBDs of the RNA-binding region of TRBP and pre-miR-155. The resulting binding isotherms were globally fitted and best described

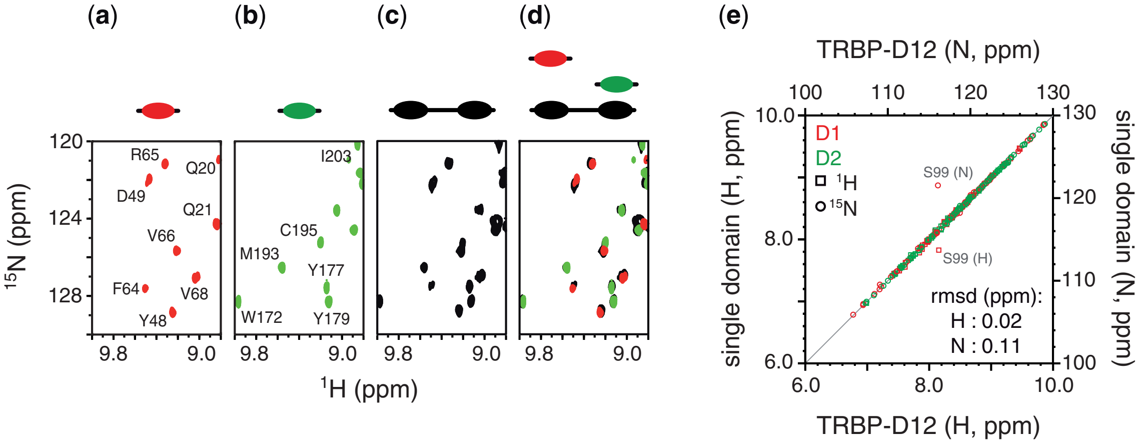


Figure 3. Comparison of NMR data collected of single- and double-domain constructs in complex with pre-miR-155. Two dimensional (^{1}H,^{15}N) spectra of (**a**) TRBP-D1/pre-miR-155 in red, (**b**) TRBP-D2/pre-miR-155 in green, and (**c**) TRBP-D12/pre-miR-155 in black. A coloured graphical representation of each construct is provided above the spectrum. (**d**) Superposition of the spectra shown in (a-c) retaining the same colour scheme and construct graphics. (**e**) Comparison of the (^{1}H,^{15}N) resonance frequencies of a residue in the single domain and the corresponding residue in TRBP-D12. Circles denote (^{15}N) chemical shifts and squares (^{1}H) chemical shifts. Amide groups from TRBP-D1 are coloured in red and those from TRBP-D2 in green. The points corresponding to the large outliers are annotated. RMSD values for ^{1}H and ^{15}N chemical shifts are provided.

by a two-site sequential binding model in which the first molecule of TRBP-D12 binds pre-miR-155 approximately 10 times more strongly than the second molecule ($K_{d,1}$ = 0.5 ± 0.03 µM; $K_{d,2}$ = 6.2 ± 0.2 µM; Figure 4a and Supplementary Table S2). The difference in affinity measured for the two molecules of TRBP-D12 is most likely due to a decrease in the number of sites available for the second binding event with respect to the first. The interaction between TRBP-D12 and pre-miR-155 exhibits an exothermic character, which suggests that polar interactions are the primary driving force of the binding event and which is consistent with the electrostatic nature of interactions between dsRNAs and dsRBDs (47).

A comparison between the binding modes of single- and double-domain constructs was also performed. ITC data obtained from the titration of TRBP-D1, TRBP-D2 or TRBP-D12 with pre-miR-155 were analysed using the standard single-site model. The resulting macroscopic dissociation constant, within the context of a identical independent sites model for the complex formed between TRBP-D12 and pre-miR-155, is 3.4 ± 0.8 µM (Supplementary Figure S4b and c and Supplementary Table S2). ITC analyses of the interactions between single-domain constructs and pre-miR-155 revealed that both single-domain constructs, TRBP-D1 and TRBP-D2, interact with pre-miR-155 to form 4:1 complexes (Supplementary Figure S4b and c and Supplementary Table S2). All macroscopic binding constants measured were in the mid-to-low micromolar range, consistent with results of TRBP-D12. TRBP-D2 (K_d = 3.7 ± 0.5 µM) interacts 4 times more strongly than TRBP-D1 (15.4 ± 0.8 µM). Like TRBP-D12, both TRBP-D1 and TRBP-D2 exhibit an exothermic binding enthalpy, again suggesting that polar interactions mediate binding. Furthermore, the macroscopic binding enthalpy obtained for TRBP-D12 is approximately equal to the sum of that of TRBP-D1 and TRBP-D2, which indicates that the polar interactions formed by each single domain and pre-miR-155 are unaffected in the context of the full-length protein. Thus, these ITC data support the conclusions derived from NMR analysis that each dsRBD interacts independently with pre-miR-155.

The formation of a complex between TRBP-D12 and pre-miR-155 in solution was evaluated by sedimentation velocity analytical ultracentrifugation (svAUC). A series of samples were prepared to characterize the behaviour of the free protein, free RNA and the protein/RNA complex under different conditions. SEC-MALLS (Figure 1b) strongly suggests that TRBP-D12 is monomeric in solution. This observation was supported by svAUC data at 10 and 20 µM, which revealed an experimental sedimentation coefficient, s_{exp}, of 2.19 S and which was independent of the protein concentration used (Figure 4b). A frictional ratio of 1.43 for TRBP-D12 was calculated by combining s_{exp} and the molecular mass (22.3 kDa) of the monomeric protein. This value is in excellent agreement with one previously reported for a protein composed of two dsRBDs connected by a flexible linker (51). When a single sedimenting species is considered, a molar mass of 22.4 kDa is calculated from the sedimentation profile of TRBP-D12, without assumptions about the shape of the sedimenting species, which is in excellent agreement with the expected molecular mass. Similar svAUC experiments indicated that pre-miR-155 is largely monomeric in solution at sample concentrations of 1 and 2 µM. A s_{exp} of 4.4 S is obtained for pre-miR-155 at both 1 µM and 2 µM, which corresponds to a frictional ratio of 1.35. Non-interacting species analysis gives a molecular mass of 18.5 kDa, which is again close to the theoretical value of 19.4 kDa. An additional minor contribution at 7.2 S (6% of the main peak) was also detected. This peak is likely to result from a small fraction of dimerized pre-miR-155. This species was also observed by SEC-MALLS (data not shown).

SvAUC analysis of the TRBP-D12/pre-miR-155 complex was performed using three samples, each with a different excess of protein. In the resulting sedimentation

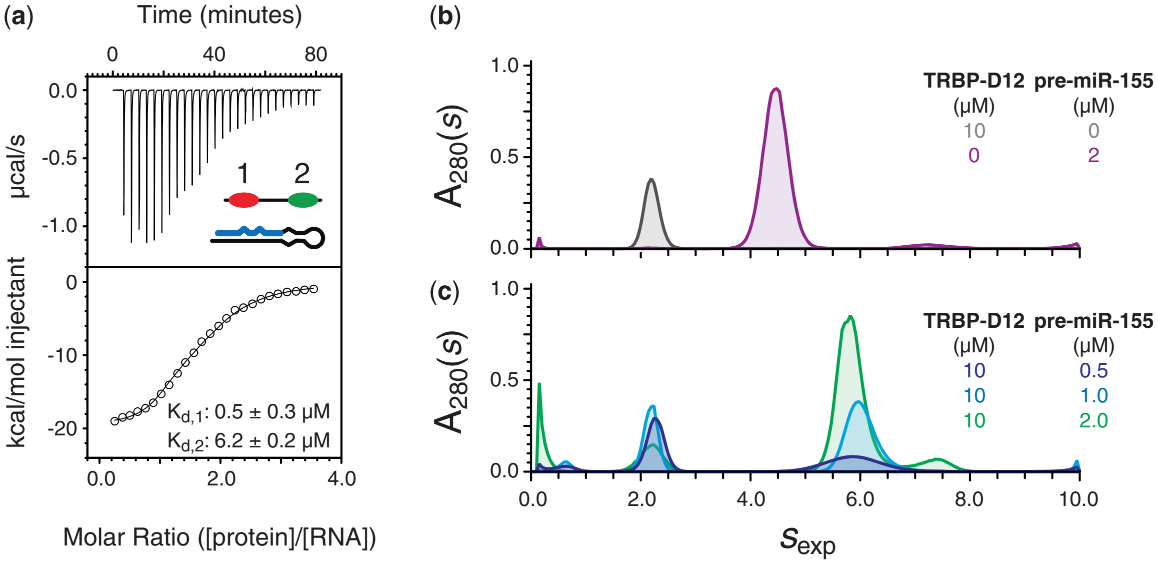


Figure 4. Characterization of the interaction between TRBP-D12 and pre-miR-155. (**a**) An example of ITC data recorded of the titration of pre-miR-155 with TRBP-D12. The top panel shows power *vs.* time together with a schematic of the components of the sample. The bottom panel shows the integrated injection enthalpy per mole of injectant plotted as a function of the ratio of protein and RNA concentrations. The data were fitted to a two-site sequential binding model shown by a solid black line. The two dissociation constants are given. The thermodynamic parameters calculated are provided in Supplementary Table S2. (**b**) Comparison of c(*s*) distributions of TRBP-D12 (grey) and pre-miR-155 (purple); (**c**) c(*s*) distributions of a sample prepared with 10 µM TRBP-D12 and different concentrations of pre-miR-155. The protein and/or RNA concentration corresponding to each curve is provided.

profiles, the RNA contribution at 4.4 S has disappeared and a new faster contribution is visible, which sediments at 5.9 S (Figure 4c). The composition of the 5.9 S peak was estimated using the data obtained from the three different optical signals. The ratios of these signals as well as theoretical values calculated for TRBP-D12/pre-miR-155 complexes with different stoichiometries are given in Supplementary Table S1. The experimental ratios measured in the c(*s*) analysis for the fast boundary (Supplementary Table S1) reflect an enrichment of protein in the complex and strongly suggest that a protein/RNA complex with a stoichiometry of 2:1 can be formed in solution.

The combination of ITC and AUC analyses demonstrates that two molecules of TRBP-D12 can interact with a single molecule of pre-miR-155. These data are consistent with ITC analysis of the two single-domain constructs, which demonstrated that four molecules of either TRBP-D1 or TRBP-D2 can interact with a single pre-miR-155. It is therefore likely that pre-miR-155 has a minimum of four dsRBD-binding sites.

The RNA-binding region of TRBP interacts with the substrates and products of Dicer

Dicer excises the apical loop of pre-miR-155 (nucleotides 24–39) to generate a largely double-stranded molecule containing the 23-nucleotide miR-155 and its complement, miR-155* (Supplementary Figure S3a). ITC experiments were conducted to characterize the interaction between miR-155/miR-155* and the RNA-binding region of TRBP (Figure 5a). These data revealed many similarities to those measured for the interaction between TRBP-D12 and pre-miR-155. The interaction between TRBP-D12 and miR-155/miR-155* was also exothermic in character (Figure 4a). In addition, the binding isotherms were also optimally described by a two-step sequential binding model ($K_{d,1}$ = 1.0 ± 0.7 µM; $K_{d,2}$ = 5.7 ± 0.3 µM; Supplementary Table S2). Finally, as was seen for the TRBP-D12/pre-miR-155 complex, the first molecule of TRBP-D12 binds miR-155/miR-155* more strongly than the second molecule. The difference in dissociation constants is again likely due to the reduced probability of the second molecule of TBRP-D12 binding.

NMR spectroscopy was used to assess which regions of TRBP-D12 interact with the miR-155/miR-155* duplex. Large chemical shift changes are observed in 2D (^{1}H,^{15}N) HSQC spectra of [^{15}N]-labelled TRBP-D12 on the addition of unlabelled miR-155/miR-155* (Figure 5b and Supplementary Figure S5b and c). Comparison of the spectra of the two TRBP-D12/RNA complexes shows little difference between the magnitude or sequence location of chemical shift perturbations. The changes in chemical shift for TRBP-D12 on binding miR-155/miR-155* are largest in the three RNA interaction regions of each dsRBD (Supplementary Figure S3b). Again, residues in the linker are barely perturbed. NMR spectra of TRBP-D12 in complex with either pre-miR-155 or miR-155/miR-155* are nearly identical (Figure 5b and Supplementary Figure S5b and c). Site-by-site differences in (^{1}H,^{15}N) resonance frequencies between single-domain and double-domain data sets are very small, with RMSD values of 0.02 and 0.11 ppm for ^{1}H and ^{15}N frequencies, respectively (Supplementary Figure S5c). These data suggest that the complexes

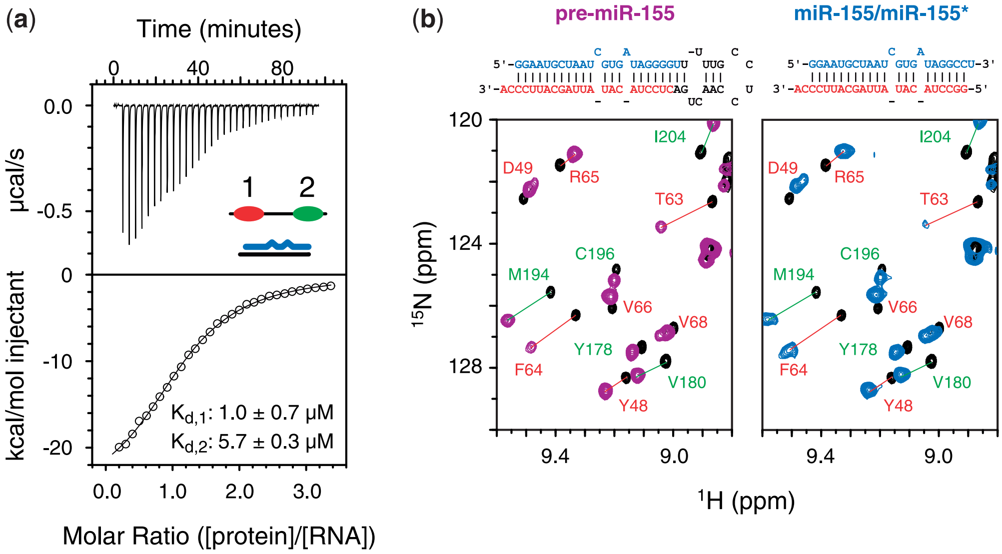


Figure 5. Interaction of the RNA-binding region of TRBP with the substrate and product of Dicer. (**a**) An example of an ITC analysis of the titration of miR-155/miR-155* with TRBP-D12 showing power *vs.* time data (top, with a schematic of the components of the sample) and the integrated injection enthalpy per mole of injectant plotted as a function of the protein/RNA ratio (bottom). The data were fitted to a two-step sequential binding model shown by a solid black line. The two dissociation constants and their standard deviations are given. The complete thermodynamic parameters derived are provided in Supplementary Table S2. (**b**) Comparison of the changes in (^{1}H,^{15}N) NMR spectra of TRBP-D12 resulting from the addition of either pre-miR-155 (left) or miR-155/miR-155* (right). The nucleotide sequence and predicted secondary structure of each RNA are provided. (^{1}H,^{15}N) assignments of cross-peaks in each spectrum are annotated and coloured according to the dsRBD in which they reside (red, dsRBD-1; green, dsRBD-2).

formed between TRBP-D12 and either pre-miR-155 or miR-155/miR-155* are highly similar.

ITC and NMR data demonstrate that TRBP-D12 interacts with two different precursors of miR-155. These data support the formation of two highly similar complexes, which are dependent on the same protein/RNA interaction surfaces. These data also suggest that the apical loop of pre-miR-155 is neither directly recognized by TRBP nor is it required for the interaction. The similarity in the measured affinities also suggests that the RNA-binding region of TRBP alone cannot differentiate between the two precursor molecules.

DISCUSSION

dsRBDs often appear in modular RNA-binding proteins, either in multiple copies or in combination with other functional domains such as RNase III, RNA helicases and PAZ domains (47). Structural studies of modular proteins containing single or multiple copies of dsRBDs have frequently revealed that these domains do not interact with other structured regions in the same protein. An NMR spectroscopic study of the dsRNA-binding region of PKR revealed no interaction between the two dsRBDs (52). Likewise, the two dsRBDs of ADAR2 (50) and HYL1 (19) were also shown to behave independently. In PKR, ADAR2 and HYL1, a long, unstructured linker connects the two dsRBDs. We have shown here that the first two dsRBDs of TRBP are also independent. Solution NMR analysis of different constructs of the RNA-binding region of TRBP revealed no contact between the two dsRBDs (Figure 1c and Supplementary Figure S1), while SEC-MALLS indicated that each isolated dsRBD is monomeric in solution.

NMR (Figure 3) and ITC analysis (Supplementary Figure S4) of TRBP-D1 and TRBP-D2 revealed that each isolated domain interacts with pre-miR-155, which is consistent with previous conclusions that these two dsRBDs constitute the main interaction site for immature miRNAs in TRBP (18,31). However, little biophysical or structural analysis of how multiple-domain constructs of TRBP interact with immature miRNAs has been previously reported. The NMR analysis of the interaction between TRBP-D12 and pre-miR-155 presented here reveals several interesting features: first, comparison of NMR spectra of single- and double-domain constructs in complex with pre-miR-155 shows no evidence of an interaction between the two dsRBDs (Figure 3). NMR spectra of each complex studied revealed no examples of peak doubling that would typically indicate the formation of multiple different protein/RNA or protein/protein interfaces. Furthermore, the very small perturbations of the (^{1}H,^{15}N) chemical shifts of the inter-domain linker indicate that this region does not interact with pre-miR-155 nor does RNA binding induce contacts with either dsRBD. The elevated signal intensity of residues in the linker with respect to the two dsRBDs is consistent with this region remaining flexible in the TRBP-D12/pre-miR-155 complex. Taken together, these data strongly suggest that each dsRBD interacts with pre-miR-155 independently. This mode of interaction is similar to that used by ADAR2. A recently elucidated 3D structure of the complex formed by ADAR2 and a 71-nt stem-loop RNA revealed that the two dsRBDs bind to distinct regions of the ds stem without any inter-domain contact or contribution from the inter-domain linker (50,53).

dsRBDs in multiple arrays do not always interact with dsRNA independently. An interesting example that contrasts TRBP is DGCR8, a component of the Microprocessor complex, which performs the first processing step in miRNA biogenesis. The two dsRBDs of DGCR8 interact with each other through a large interface, which fixes their relative orientation and has important implications for pri-miRNA recognition. DGCR8 interacts with pri-miRNAs in preference to pre-miRNAs (54,55), whereas our data show that TRBP can interact with both pre-miRNAs and miRNA/miRNA* duplexes with similar affinities. These different patterns of RNA selectivity displayed by DGCR8 and TRBP could result from the contrasting ways in which the tandem dsRBDs interact with immature miRNAs. The compact structure formed by the DGCR8 core means that the two dsRBDs are not independent and that a pri-miRNA must bend to interact with both RNA-binding surfaces (56). A sufficient conformational change may not be feasible for the shorter pre-miRNA, leading to the preference for pri-miRNAs. Conversely, given that TRBP has been implicated in both the Dicer cleavage reaction and the transfer of the miRNA/miRNA* duplex to RISC, the independence of its two dsRBDs demonstrated here may reflect a broader range of binding targets. Taken together, these results suggest that there are considerable differences in how RNA is recognized by the different accessory proteins involved in the two miRNA biogenesis steps.

A second important feature of the TRBP-D12/pre-miR-155 complex is the stoichiometry. Both ITC (Figure 4a and Supplementary Figure S4) and svAUC (Figure 4b) experiments indicate that a complex containing four dsRBDs can be formed; that is, a single pre-miR-155 interacts with four single dsRBD constructs or two double-domain constructs. ITC data of the interactions between TRBP-D12 and pre-miR-155, and between TRBP-D12 and miR155/miR155*, revealed no evidence of positive co-operativity. This result is consistent with earlier studies (34).

The data presented here indicate that one pre-miR-155 can associate with four dsRBDs to form a complex in which the dsRBDs interact with similar sites on the RNA but do not form inter-domain contacts. Is there sufficient space on pre-miR-155 to permit this? The theoretical secondary structure of pre-miR-155 is typical of pre-miRNAs and comprises an imperfect ds stem containing two GU base-pairs, two unpaired nucleotides and a four-base apical loop (Figure 5b). The majority of pre-miRNAs in the miRbase (23) are predicted to adopt 3D structures similar to A-form dsRNA (57). Three-dimensional structures of large stem-loop RNAs have revealed that features such as non-Watson–Crick base-pairs or nucleotide bulges often cause minimal perturbation to the overall ds character (53,58). Canonical dsRBDs typically interact with a 12–16-bp region of dsRNA (47). However, the RNA-binding region of PKR, which also contains two dsRBDs, can interact with 20–40-bp RNA duplexes with stoichiometries that are incompatible with a simple 1D lattice model and which would require binding to multiple surfaces of the dsRNA with a binding site overlap of 4 bp (51). Several 3D structures of protein/dsRNA complexes have also demonstrated that multiple dsRBDs can bind in proximity without forming inter-dsRBD interactions (53,59).

A structural model of the interaction between pre-miR-155 and dsRBDs was generated to evaluate the feasibility of a protein/RNA complex containing the four dsRBDs per pre-miR-155 suggested by ITC and AUC analysis (Supplementary Figure S6). The model demonstrates that it is theoretically possible to arrange four non-contacting dsRBDs on a single molecule of pre-miR-155. The dsRBDs tessellate optimally when they interact with the RNA in alternating directions. The inter-dsRBD linker in TRBP is more than 60 residues and therefore places little limitation on the respective orientation of the two dsRBDs in the TRBP-D12 construct. In addition, NMR data suggest that the linker remains flexible in both of the TRBP-D12/RNA complexes analysed here. The arrangement shown in Supplementary Figure S6 is also compatible with NMR and ITC data of the interaction between TRBP-D12 and the Dicer product miR-155/miR-155*, as none of the four proposed dsRBD-binding sites in the model involve nucleotides from the excised apical loop.

The functional stoichiometry of the Dicer/TRBP complex has not been determined (32). However, the report that a tertiary complex containing Dicer, TRBP and PACT can be formed implies that two TRBP-like proteins can interact simultaneously with Dicer in the presence of pre-miRNAs (21). In addition, the plant DCL1 accessory protein HYL1 has been shown to form 2:1 protein/RNA complexes with dsRNA sequences of suitable length (18). Thus, high occupancy of dsRBDs on a relatively short sequence of dsRNA, as seen here with TRBP and two different precursors of miR-155, is potentially functionally relevant. Whether the 2:1 ratio observed here is a standard feature of the complexes formed between TRBP and immature miRNAs or whether the stoichiometry is determined by features in the target immature miRNA, such as the length or percentage of canonical dsRNA, remains to be determined.

Much of the information concerning the role of TRBP in small regulatory RNA synthesis has been obtained from studies into siRNAs. The data presented here represent the first in-depth biophysical analysis of the interaction between TRBP and immature miRNAs that have an imperfect dsRNA stem. Analysis of ITC data yielded macroscopic dissociation constants for the TRBP/pre-miR-155 complex in the low µM range, which is also consistent with other investigations of dsRBD/dsRNA interactions (18,19,56,60,61). The affinities measured for the isolated dsRBDs of TRBP indicate that TRBP-D2 interacts with pre-miR-155 approximately 4 times more strongly than TRBP-D1. This observation is consistent with earlier interaction studies of TRBP and siRNAs (31). Differences between the binding affinities of dsRBDs in tandem or multi-dsRBD proteins are relatively common—e.g. HYL1 (18,19) or PKR (45)—and may play a role in target selection by fine-tuning the overall affinity of a protein for dsRNA molecules that possess many potential dsRBD-binding sites.

In addition to its function in the Dicer complex, TRBP is a component of the RLC (4,11,12). The precise role of TRBP in the formation and/or activity of RISC is less clear. We have shown here that the RNA-binding region of TRBP interacts with both pre-miR-155 and miR-155/miR-155*—i.e. both the substrate and product of Dicer—with a comparable affinity (Figure 5a). Furthermore, NMR spectra of TRBP-D12 in complex with either pre-miR-155 or miR-155/miR-155* are highly similar (Figure 5b), which indicates that TRBP interacts with pre-miR-155 and miR-155/miR-155* through essentially identical sites. These results suggest that TRBP could participate in the recognition of Dicer substrates as well as the transfer of the cleavage product to RISC through the RLC.

CONCLUSIONS

Despite several structural studies, many of the mechanistic details that underpin human miRNA biogenesis remain unclear. One area in which the link between 3D structure and function remains particularly obscure concerns the role of the non-catalytic proteins of that associate with Drosha or Dicer. We have demonstrated here that the region of TRBP responsible for interacting with immature miRNAs is composed of two dsRBDs that recognize RNA independently. The linker connecting the two dsRBDs in the RNA-binding region of TRBP remains flexible and does not interact with the bound RNA. Therefore, the formation of a complex between TRBP and immature miRNAs is unlikely to require bending of the ds stem of the RNA, as has been reported for DGCR8. Finally, the observation that RNA-binding region of TRBP can interact with both the substrates and products of Dicer suggests that TRBP may function at several stages in the miRNA biogenesis pathway.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2, Supplementary Figures 1–6, Supplementary Methods and Supplementary References [62–64].

ACKNOWLEDGEMENTS

The authors thank Drs. A. Favier and E. Condamine for help with the acquisition of NMR data; Drs. R. Rasia, B. Bersch, C. Petosa and A. Leech for critical reading of the manuscript; and the Grenoble Partnership for Structural Biology (PSB) for access to integrated structural biology platforms, including: A. le Roy from the IBS/PSB AUC platform for help in the acquisition and analysis of AUC data; Dr. M. Noirclerc-Savoye, from the IBS/PSB RoBioMol platform, for the generation of the constructs used in this study; and the IBS/PSB high-field NMR spectroscopy platform; and IBS/PSB cell free expression platform.

FUNDING

This work was supported by grants from the European Marie Curie Programme (International Reintegration Grant 231082 to M.J.P.) and l'Agence Nationale de la Recherche (06-JCJC-0034 to J.B.). M.J.P. acknowledges a post-doctoral fellowship from l'Association pour la Recherche sur le Cancer, France, and M.P.M.H.B. acknowledges a Ph.D. fellowship from le Commissariat à l'Énergie Atomique et aux Energies Alternatives (CEA), France. Funding for open access charge: CNRS.

Conflict of interest statement. None declared.

REFERENCES

1. Zamore,P.D. and Haley,B. (2005) Ribo-gnome: the big world of small RNAs. *Science*, **309**, 1519–1524.
2. Liu,Q. and Paroo,Z. (2010) Biochemical principles of small RNA pathways. *Annu. Rev. Biochem.*, **79**, 295–319.
3. Haase,A.D., Jaskiewicz,L., Zhang,H., Lainé,S., Sack,R., Gatignol,A. and Filipowicz,W. (2005) TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep.*, **6**, 961–967.
4. Chendrimada,T.P., Gregory,R.I., Kumaraswamy,E., Norman,J., Cooch,N., Nishikura,K. and Shiekhattar,R. (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, **436**, 740–744.
5. Gatignol,A., Buckler-White,A., Berkhout,B. and Jeang,K.T. (1991) Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. *Science*, **251**, 1597–1600.
6. Park,H., Davies,M.V., Langland,J.O., Chang,H.W., Nam,Y.S., Tartaglia,J., Paoletti,E., Jacobs,B.L., Kaufman,R.J. and Venkatesan,S. (1994) TAR RNA-binding protein is an inhibitor of the interferon-induced protein kinase PKR. *Proc. Natl Acad. Sci. U.S.A.*, **91**, 4713–4717.
7. Cosentino,G.P., Venkatesan,S., Serluca,F.C., Green,S.R., Mathews,M.B. and Sonenberg,N. (1995) Double-stranded-RNA-dependent protein kinase and TAR RNA-binding protein form homo- and heterodimers in vivo. *Proc. Natl Acad. Sci. USA.*, **92**, 9445–9449.
8. Hutvágner,G., McLachlan,J., Pasquinelli,A.E., Bálint,E., Tuschl,T. and Zamore,P.D. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science*, **293**, 834–838.
9. Bernstein,E., Caudy,A.A., Hammond,S.M. and Hannon,G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, **409**, 363–366.
10. Chakravarthy,S., Sternberg,S.H., Kellenberger,C.A. and Doudna,J.A. (2010) Substrate-specific kinetics of Dicer-catalyzed RNA processing. *J. Mol. Biol.*, **404**, 392–402.
11. Gregory,R.I., Chendrimada,T.P., Cooch,N. and Shiekhattar,R. (2005) Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell*, **123**, 631–640.
12. MacRae,I.J., Ma,E., Zhou,M., Robinson,C.V. and Doudna,J.A. (2008) In vitro reconstitution of the human RISC-loading complex. *Proc. Natl Acad. Sci. USA.*, **105**, 512–517.
13. Liu,Q., Rand,T.A., Kalidas,S., Du,F., Kim,H.E., Smith,D.P. and Wang,X. (2003) R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. *Science*, **301**, 1921–1925.
14. Förstemann,K., Tomari,Y., Du,T., Vagin,V.V., Denli,A.M., Bratu,D.P., Klattenhoff,C., Theurkauf,W.E. and Zamore,P.D. (2005) Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.*, **3**, e236.
15. Saito,K., Ishizuka,A., Siomi,H. and Siomi,M.C. (2005) Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. *PLoS Biol.*, **3**, e235.
16. Tabara,H., Yigit,E., Siomi,H. and Mello,C.C. (2002) The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a

DExH-box helicase to direct RNAi in C elegans. *Cell*, **109**, 861–871.
17. Vazquez,F., Gasciolli,V., Crété,P. and Vaucheret,H. (2004) The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. *Curr. Biol.*, **14**, 346–351.
18. Yang,S.W., Chen,H.Y., Yang,J., Machida,S., Chua,N.H. and Yuan,Y.A. (2010) Structure of Arabidopsis HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. *Structure*, **18**, 594–605.
19. Rasia,R.M., Mateos,J., Bologna,N.G., Burdisso,P., Imbert,L., Palatnik,J.F. and Boisbouvier,J. (2010) Structure and RNA interactions of the plant MicroRNA processing-associated protein HYL1. *Biochemistry*, **49**, 8237–8239.
20. Lee,Y., Hur,I., Park,S.Y., Kim,Y.K., Suh,M.R. and Kim,V.N. (2006) The role of PACT in the RNA silencing pathway. *EMBO J.*, **25**, 522–532.
21. Kok,K.H., Ng,M.H.J., Ching,Y.P. and Jin,D.Y. (2007) Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA. *J. Biol. Chem.*, **282**, 17649–17657.
22. Patel,R.C. and Sen,G.C. (1998) PACT, a protein activator of the interferon-induced protein kinase, PKR. *EMBO J.*, **17**, 4379–4390.
23. Griffiths-Jones,S., Grocock,R.J., van Dongen,S., Bateman,A. and Enright,A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.
24. Garzon,R., Calin,G.A. and Croce,C.M. (2009) MicroRNAs in Cancer. *Annu. Rev. Med.*, **60**, 167–179.
25. Negrini,M., Nicoloso,M.S. and Calin,G.A. (2009) MicroRNAs and cancer—new paradigms in molecular oncology. *Curr. Opin. Cell Biol.*, **21**, 470–479.
26. Melo,S.A., Ropero,S., Moutinho,C., Aaltonen,L.A., Yamamoto,H., Calin,G.A., Rossi,S., Fernandez,A.F., Carneiro,F., Oliveira,C. *et al.* (2009) A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. *Nat. Genet.*, **41**, 365–370.
27. Paroo,Z., Ye,X., Chen,S. and Liu,Q. (2009) Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling. *Cell*, **139**, 112–122.
28. Shan,G., Li,Y., Zhang,J., Li,W., Szulwach,K.E., Duan,R., Faghihi,M.A., Khalil,A.M., Lu,L., Paroo,Z. *et al.* (2008) A small molecule enhances RNA interference and promotes microRNA processing. *Nat. Biotechnol.*, **26**, 933–940.
29. Jinek,M. and Doudna,J.A. (2009) A three-dimensional view of the molecular machinery of RNA interference. *Nature*, **457**, 405–412.
30. Parker,J.S. (2010) How to slice: snapshots of Argonaute in action. *Silence*, **1**, 3.
31. Yamashita,S., Nagata,T., Kawazoe,M., Takemoto,C., Kigawa,T., Güntert,P., Kobayashi,N., Terada,T., Shirouzu,M., Wakiyama,M. *et al.* (2011) Structures of the first and second double-stranded RNA-binding domains of human TAR RNA-binding protein. *Protein Sci.*, **20**, 118–130.
32. Lau,P.W., Potter,C.S., Carragher,B. and MacRae,I.J. (2009) Structure of the human Dicer-TRBP complex by electron microscopy. *Structure*, **17**, 1326–1332.
33. Wang,H.W., Noland,C., Siridechadilok,B., Taylor,D.W., Ma,E., Felderer,K., Doudna,J.A. and Nogales,E. (2009) Structural insights into RNA processing by the human RISC-loading complex. *Nat. Struct. Mol. Biol.*, **16**, 1148–1153.
34. Parker,G.S., Maity,T.S. and Bass,B.L. (2008) dsRNA binding properties of RDE-4 and TRBP reflect their distinct roles in RNAi. *J. Mol. Biol.*, **384**, 967–979.
35. Benoit,M.P. and Plevin,M.J. (2013) Backbone resonance assignments of the micro-RNA precursor binding region of human TRBP. *Biomol. NMR Assign.* August 9 (doi:10.1007/s12104-012-9416-8; epub ahead of print).
36. Lescop,E., Schanda,P. and Brutscher,B. (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.*, **187**, 163–169.
37. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
38. Vranken,W.F., Boucher,W., Stevens,T.J., Fogh,R.H., Pajon,A., Llinas,M., Ulrich,E.L., Markley,J.L., Ionides,J. and Laue,E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
39. Mulder,F.A., Schipper,D., Bott,R. and Boelens,R. (1999) Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins. *J. Mol. Biol.*, **292**, 111–123.
40. Laue,T.M., Shah,B.D., Ridgeway,T.M. and Pelletier,S.L.Computer-aided interpretation of analytical sedimentation data for proteins. In: Horton,J., Rowe,A. and Harding,S.E. (eds), *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. Royal Society of Chemistry, Cambridge, pp. 90–125.
41. Bonifacio,G.F., Brown,T., Conn,G.L. and Lane,A.N. (1997) Comparison of the electrophoretic and hydrodynamic properties of DNA and RNA oligonucleotide duplexes. *Biophys. J.*, **73**, 1532–1538.
42. Schuck,P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.*, **78**, 1606–1619.
43. Balbo,A., Minor,K.H., Velikovsky,C.A., Mariuzza,R.A., Peterson,C.B. and Schuck,P. (2005) Studying multiprotein complexes by multisignal sedimentation velocity analytical ultracentrifugation. *Proc. Natl Acad. Sci. USA.*, **102**, 81–86.
44. Keller,S., Vargas,C., Zhao,H., Piszczek,G., Brautigam,C.A. and Schuck,P. (2012) High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal. Chem.*, **84**, 5066–5073.
45. Nanduri,S., Rahman,F., Williams,B.R. and Qin,J. (2000) A dynamically tuned double-stranded RNA binding mechanism for the activation of antiviral kinase PKR. *EMBO J.*, **19**, 5567–5574.
46. Cavallo,L., Kleinjung,J. and Fraternali,F. (2003) POPS: a fast algorithm for solvent accessible surface areas at atomic and residue level. *Nucleic Acids Res.*, **31**, 3364–3366.
47. Tian,B., Bevilacqua,P.C., Diegelman-Parente,A. and Mathews,M.B. (2004) The double-stranded-RNA-binding motif: interference and much more. *Nat. Rev. Mol. Cell Biol.*, **5**, 1013–1023.
48. Yang,X., Welch,J.L., Arnold,J.J. and Boehr,D.D. (2010) Long-range interaction networks in the function and fidelity of poliovirus RNA-dependent RNA polymerase studied by nuclear magnetic resonance. *Biochemistry*, **49**, 9361–9371.
49. Wang,Z., Hartman,E., Roy,K., Chanfreau,G. and Feigon,J. (2011) Structure of a yeast RNase III dsRBD complex with a noncanonical RNA substrate provides new insights into binding specificity of dsRBDs. *Structure*, **19**, 999–1010.
50. Stefl,R., Xu,M., Skrisovska,L., Emeson,R.B. and Allain,F.H.T. (2006) Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs. *Structure*, **14**, 345–355.
51. Ucci,J.W., Kobayashi,Y., Choi,G., Alexandrescu,A.T. and Cole,J.L. (2007) Mechanism of interaction of the double-stranded RNA (dsRNA) binding domain of protein kinase R with short dsRNA sequences. *Biochemistry*, **46**, 55–65.
52. Nanduri,S., Carpick,B.W., Yang,Y., Williams,B.R. and Qin,J. (1998) Structure of the double-stranded RNA-binding domain of the protein kinase PKR reveals the molecular basis of its dsRNA-mediated activation. *EMBO J.*, **17**, 5458–5465.
53. Stefl,R., Oberstrass,F.C., Hood,J.L., Jourdan,M., Zimmermann,M., Skrisovska,L., Maris,C., Peng,L., Hofr,C., Emeson,R.B. *et al.* (2010) The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell*, **143**, 225–237.
54. Han,J., Lee,Y., Yeom,K.H., Nam,J.W., Heo,I., Rhee,J.K., Sohn,S.Y., Cho,Y., Zhang,B.T. and Kim,V.N. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, **125**, 887–901.
55. Faller,M., Toso,D., Matsunaga,M., Atanasov,I., Senturia,R., Chen,Y., Zhou,Z.H. and Guo,F. (2010) DGCR8 recognizes primary transcripts of microRNAs through highly cooperative binding and formation of higher-order structures. *RNA*, **16**, 1570–1583.
56. Sohn,S.Y., Bae,W.J., Kim,J.J., Yeom,K.H., Kim,V.N. and Cho,Y. (2007) Crystal structure of human DGCR8 core. *Nat. Struct. Mol. Biol.*, **14**, 847–853.

57. Parisien,M. and Major,F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51–55.
58. Nomura,Y., Kajikawa,M., Baba,S., Nakazato,S., Imai,T., Sakamoto,T., Okada,N. and Kawai,G. (2006) Solution structure and functional importance of a conserved RNA hairpin of eel LINE UnaL2. *Nucleic Acids Res.*, **34**, 5184–5193.
59. Gan,J., Tropea,J.E., Austin,B.P., Court,D.L., Waugh,D.S. and Ji,X. (2006) Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell*, **124**, 355–366.
60. Dong,Z., Han,M.H. and Fedoroff,N. (2008) The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. *Proc. Natl Acad. Sci. USA*, **105**, 9970–9975.
61. Wostenberg,C., Quarles,K.A. and Showalter,S.A. (2010) Dynamic origins of differential RNA binding function in two dsRBDs from the miRNA "microprocessor" complex. *Biochemistry*, **49**, 10728–10736.
62. Akaike,H. (1973) A new look at statistical model identification. *IEEE Trans. Automat. Contr.*, **19**, 716–723.
63. Burnham,K.P. and Anderson,D.R. (2004) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn. Springer.
64. Das,R., Karanicolas,J. and Baker,D. (2010) Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods*, **7**, 291–294.

The RNA-Binding Region of Human TRBP Interacts with MicroRNA precursors via Two Independent Domains

Matthieu P. M. H. Benoit, Lionel Imbert, Andrés Palencia, Julien Pérard, Christine Ebel, Jérôme Boisbouvier and Michael J. Plevin*

Supplementary Information

Multi-signal analysis of SvAUC data.

Sedimentation profiles were obtained using three optical signals and were analyzed simultaneously with the program SEDPHAT (1) using a hybrid model of one non-interacting species (for either TRBP-D12 or pre-miR-155) and c(*s*) distributions representing minor species (e.g. an RNA dimer). Data sets concerning TRBP-D12 and pre-miR-155 were analysed separately.

The extinction coefficient at 280 nm for TRBP-D12 (10,084 cm^{-1} mol^{-1}) was fixed in order to derive an extinction coefficient at 260 nm (5,814 cm^{-1} mol^{-1}) and the fringe shift (ΔJ) increment (57,569 fringe cm^{-1} mol^{-1}) from the peak integrals. The extinction coefficient at 260 nm for pre-miR-155 (490,000 cm^{-1} mol^{-1}) was fixed in order to allow the extinction coefficient at 280 nm (225,975 cm^{-1} mol^{-1}) and the fringe shift increment (69,347 fringe cm^{-1} mol^{-1}) to be derived. Integration of the fast peak of the c(*s*) curves was used to determine the experimental ratios for the fast boundary.

Comparison of ITC binding models

ITC experiments in which pre-miR-155 was titrated into TRBP-D12 suggested the formation of complex in which more than one molecule of TRBP-D12 can bind pre-miR-155 (see below). Titrations of RNA samples (either pre-miR-155 or miR-155/miR-155*) with TRBP-D12 were consequently analysed using different models, including examples with successive binding of multiple molecules of TRBP-D12 to each miR-155 precursor. Only two protein/RNA interaction models could be reliably fitted to these data: a simple 1:1 model and a successive binding model with 2:1 stoichiometry. These two models are compared using Akaike weights (2, 3) and RMSDs in the table below. Both Akaike weights and RMSDs show that models in which two molecules of protein bind successively to a single molecule of RNA (either pre-miR-155 or miR-155/miR-155*) best explain the experiment data.

System	Stoichiometry		RMSD (kcal/mol)	Akaike weights
	Protein	RNA		
TRBP-D12 + pre-miR-155	1	1	0.13	1×10^{-28}
TRBP-D12 + pre-miR-155	2	1	0.05	$\simeq 1$
TRBP-D12 + miR-155/miR-155*	1	1	0.14	1×10^{-31}
TRBP-D12 + miR-155/miR-155*	2	1	0.04	$\simeq 1$

ITC experiments performed using single dsRBD constructs (e.g. TRBP-D1 and TRBP-D2) yielded binding isotherms with a single visible step, which were best described by a complex stoichiometry of 4 protein molecules per RNA. These data do not provide sufficient constraints to allow the fitting of the 4 microscopic binding constants and enthalpies. Therefore, a single-site model was used to analyze these data. While a single-site model is not strictly appropriate for describing such interactions, it does permit a global comparison of the binding parameters that describe the interaction between the three TRBP constructs and pre-miR-155.

Modelling of the TRBP-D12/miRNA precursor complex

The 3D model of pre-miR-155 presented in Supplementary Figure 6 was generated using Rosetta (4). The 61-nucleotide sequence of pre-miR-155 (including the modifications made to the native sequence described) was split into three regions: two from the double stranded stem (region 1: 1-11 and 49-61; and region 2: 9-22 and 40-51) and one encompassing the apical loop (region 3: 19-43). Ensembles of 20 structures of the 3 fragments were generated using Rosetta (http://rosettaserver.graylab.jhu.edu). A pair-wise structural superposition of overlapping sequences was performed using ProFit (www.bioinf.org.uk/software/profit/index.html). Each aligned pair of structures were ranked. In each instance the structure from region 2 was defined as the template. A single 3D model of pre-miR-155 was constructed by combining aligned pairs of fragments with the same central region (region 2) with the lowest combined rank.

The model of the interaction between dsRBDs and pre-miR-155 was generated using the structure of TRBP-D2 in complex with a short 10 bp dsRNA (PDB code: 3ADL; (5)). The RNA fragment in 3ADL was extended by adding a symmetry-related dsRNA molecule to generate a 20 bp broken dsRNA fragment. The 20 bp dsRNA in the modified 3ADL was structurally aligned to the Rosetta model of pre-miR-155 using ProFit. The superposition was only performed for carbon atoms in the ribose ring of the template and mobile structures over a 16 bp window. This alignment process generated a model with a dsRBD molecule located at a potential binding site on the surface of pre-miR-155. The alignment was repeated by incrementing the starting position of the alignment window along the pre-miR-155 sequence to generate 28 models with a rmsd for like ribose atoms of less than 4.5 Å. A relatively high cut-off was selected to take account of the presence non-base-paired nucleotides (12, 16 and 37) and non-Watson Crick base pairing (21 and 41; and 24 and 38) in certain alignment windows.

Whilst clearly theoretical, this model does suggest that it is feasible for 4 dsRBDs to interact with both pre-miR-155 and miR-155/miR-155* without protein-protein contact.

Supplementary References

1. Balbo,A., Minor,K.H., Velikovsky,C.A., Mariuzza,R.A., Peterson,C.B. and Schuck,P. (2005) Studying multiprotein complexes by multisignal sedimentation velocity analytical ultracentrifugation. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 81–86.

2. Akaike,H. (1973) A new look at statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

3. Burnham,K.P. and Anderson,D.R. (2004) Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach 2nd ed. Springer.

4. Das,R., Karanicolas,J. and Baker,D. (2010) Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods*, **7**, 291–294.

5. Yang,S.W., Chen,H.-Y., Yang,J., Machida,S., Chua,N.-H. and Yuan,Y.A. (2010) Structure of Arabidopsis HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. *Structure*, **18**, 594–605.

Supplementary Figure Legends

Supplementary Figure S1. Comparison of NMR data collected on different constructs of apo-TRBP. 2D (^{1}H,^{15}N) spectra of (a) apo-TRBP-D1 in red, (b) apo-TRBP-D2 in green, and (c) apo-TRBP-D12 black. A colored graphical representation of each construct is provided above the spectrum. (d) superposition of the spectra shown in (a-c) retaining the same color scheme and construct graphics. The spectrum of apo-TRBP-D12 can be readily recapitulated by superposing spectra of apo-TRBP-D1 and apo-TRBP-D2. Additional peaks not present in the red and green spectra originate from residues found in the interdomain linker in TRBP-D12. (e) Compound chemical shift differences between respective sites in single (either TBRP-D1 or TRBP-D2) and double domain (i.e. TRBP-D12) constructs plotted as a function of residue number. The large deviation seen at the C-terminus of TRBP-D1 is due to the presence of the inter-domain linker in TRBP-D12.

Supplementary Figure S2. Site-by-site comparison of the (^{1}H,^{15}N) chemical shift of a residue in the single domain and the corresponding residue in TRBP-D12. Circles denote (^{15}N) chemical shifts and squares (^{1}H) chemical shifts. Rmsd values for ^{1}H and ^{15}N chemical shifts are provided.

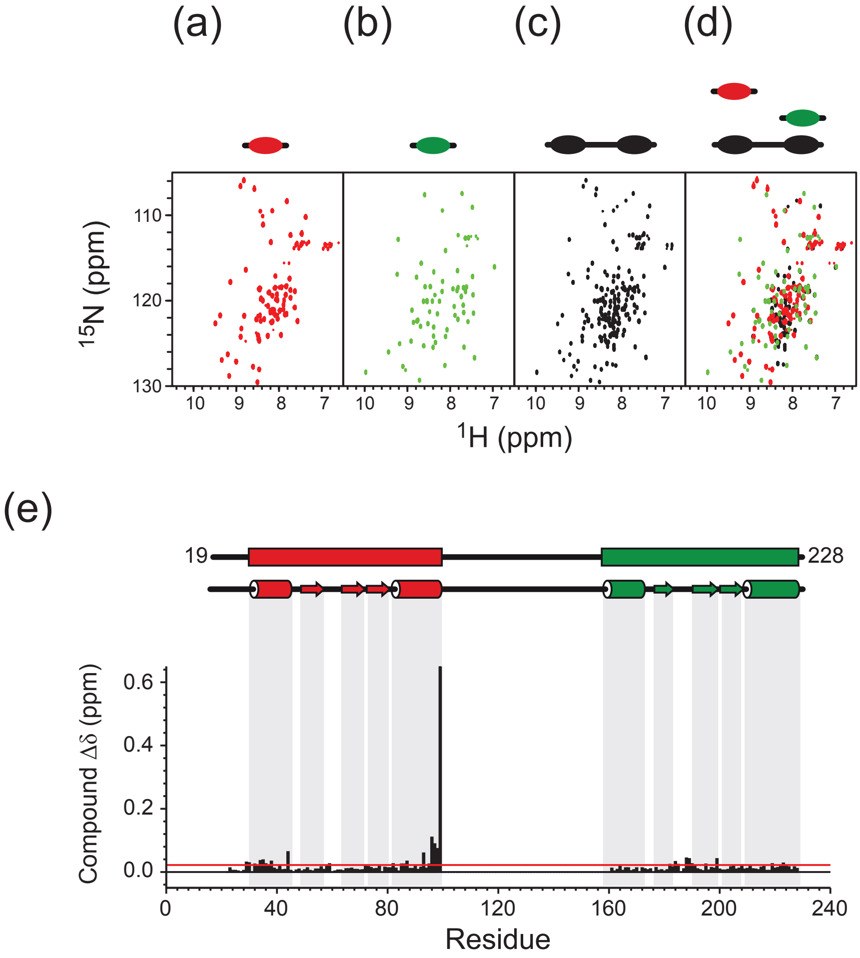
Supplementary Figure S3. Examples of the changes observed in peak position following the titration of TRBP-D12 with pre-miR-155. Regions of the spectra corresponding to L51, H58 and H85 from the first dsRBD, and R189, T208 and K211 from the second dsRBD are shown. The spectra are colored according the to RNA/protein ratio as defined in the inset.

Supplementary Figure S4. ITC analysis of the interaction between pre-miR-155 and single or double domain constructs. Examples of ITC data recorded when pre-miR-155 was titrated into (a) TRBP-D12, (b) TRBP-D1 and (c) TRBP-D2 constructs. Top panels show power *vs.* time. A schematic representation of the macromolecules in the sample is provided. The bottom panels show the integrated injection enthalpy per mole of injectant plotted as a function of the ratio of the protein and pre-miR-155 concentrations. The fit of each titration curve to a single-site binding model is shown by a solid red line. The parameters calculated for each fit are provided in the Supplementary Table S2.

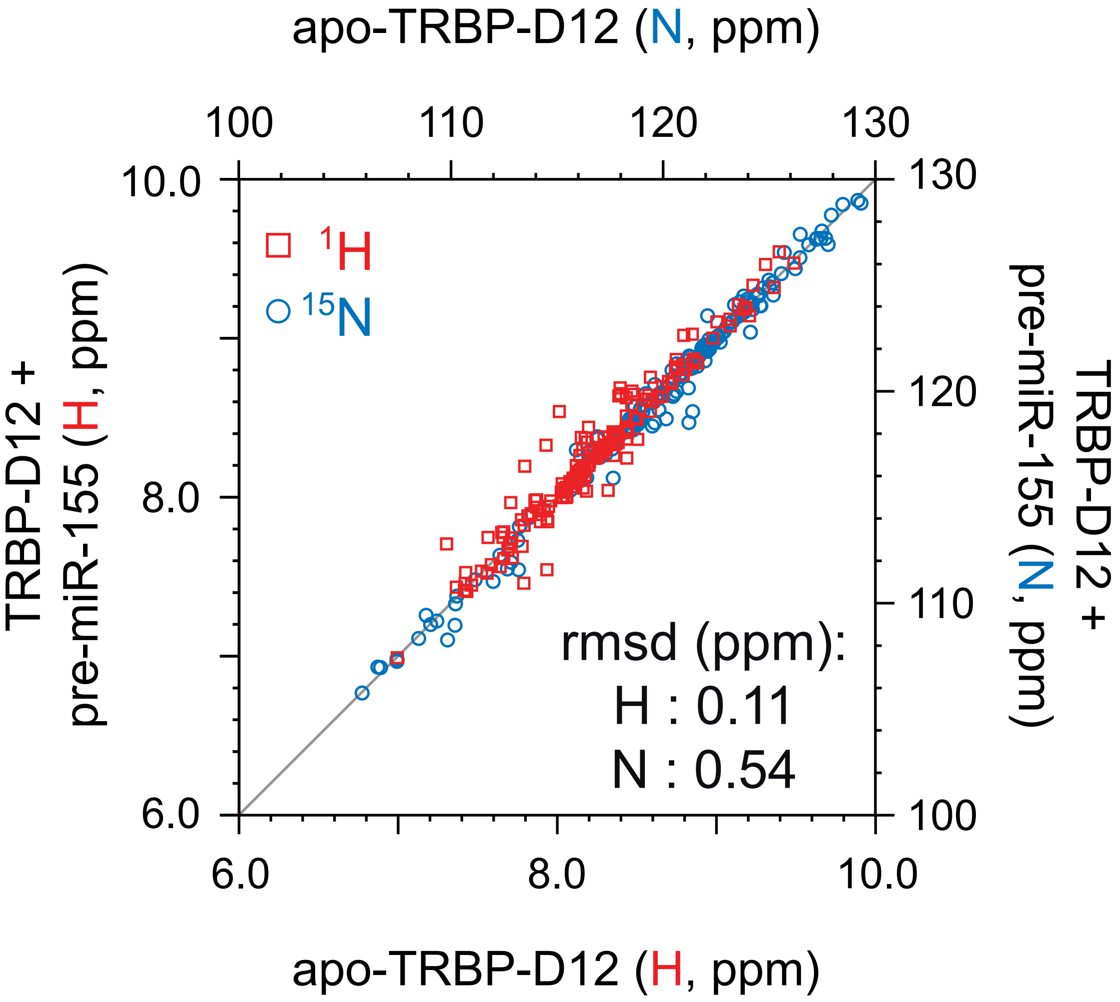
Supplementary Figure S5. (a) graphical depiction of the reaction catalysed by Dicer. The apical loop of re-miR-155 is removed following cleavage at the two sites indicated by yellow triangles. The nucleotide sequence and predicted secondary structure of pre-miR-155 and miR-155/miR-155* are shown; (b) plot of the compound chemical shift changes, $\Delta\delta$, recorded when TRBP-D12 is mixed with pre-miR-155 (purple) or miR-155/miR-155* (black). The location of domain boundaries, secondary structure elements and the 3 canonical dsRBD sites implicated in dsRNA binding are shown. (c) site-by-site comparison of the (^{1}H,^{15}N) chemical shift of a residue in the single domain and the corresponding residue in TRBP-D12. Circles denote (^{15}N) chemical shifts and squares (^{1}H) chemical shifts. The points corresponding to the large outliers are annotated in grey. Rmsd values for ^{1}H and ^{15}N chemical shifts are provided.

Supplementary Figure S6.
3D models of the interaction of dsRBDs from TRBP with pre-miR-155 (left) and miR-155/miR-155* (right). The dsRNA is coloured grey and shown as a surface representation. The 4 dsRBD molecules are coloured in different shades of blue or purple and shown either as surface (top) or cartoon (bottom) representations. The 3 canonical dsRBD interaction sites are indicated in the bottom left-hand most figure.

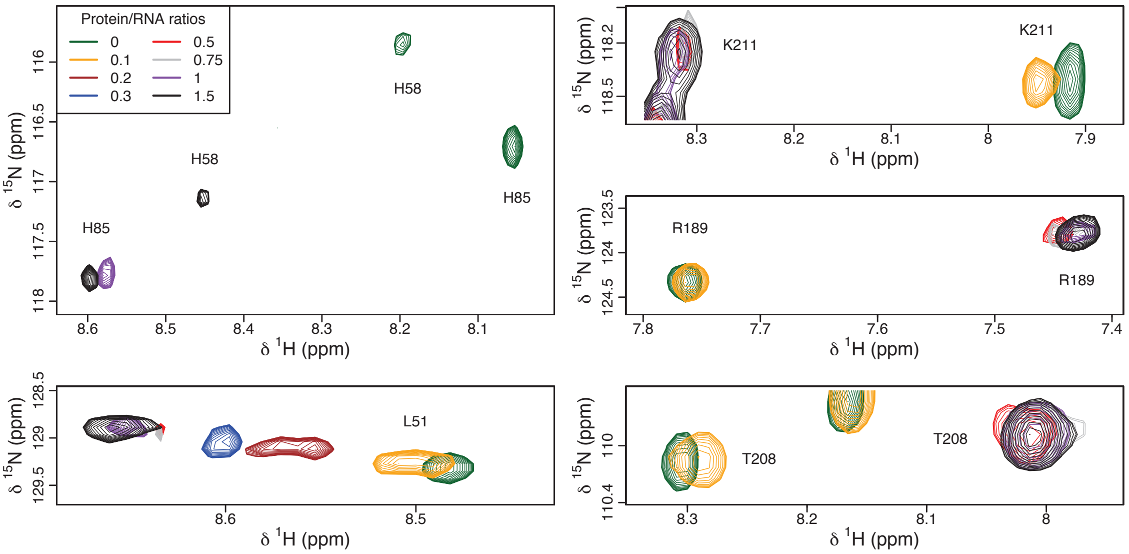
Supplementary Figure S1



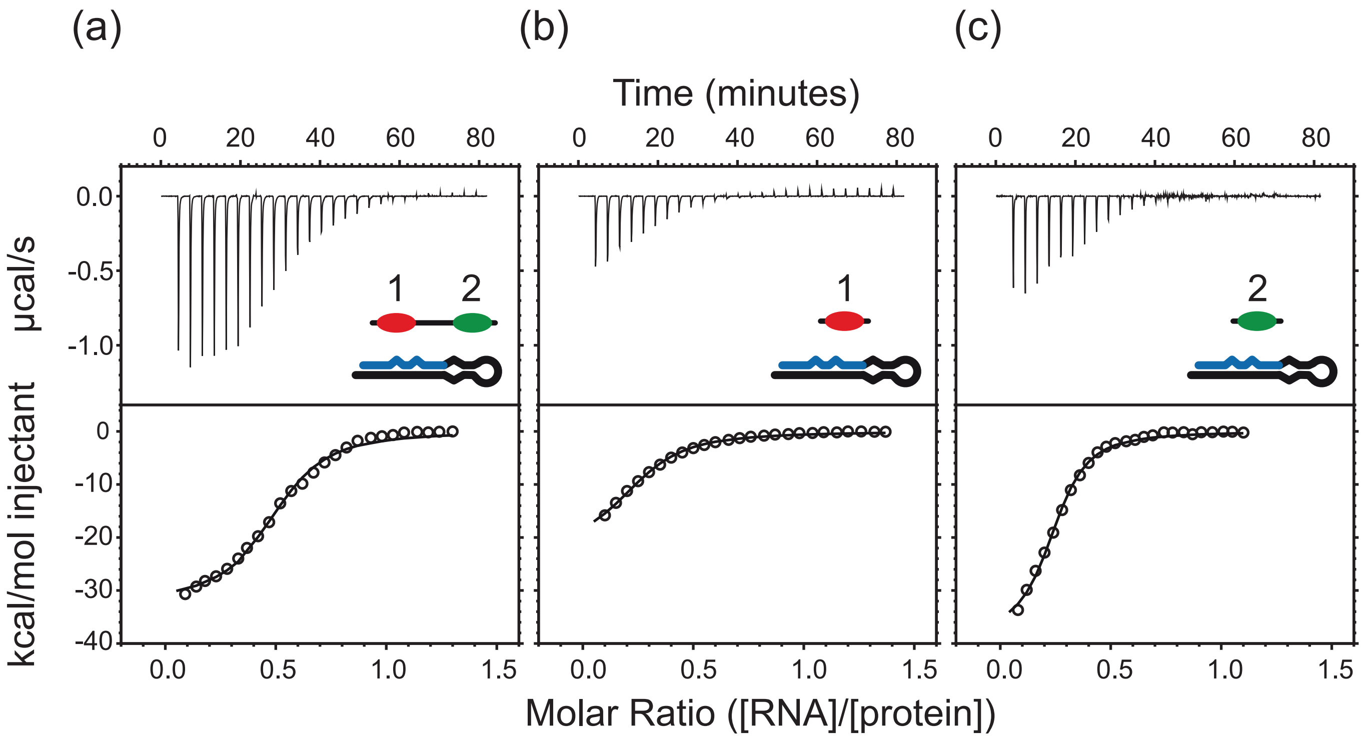
Supplementary Figure S2



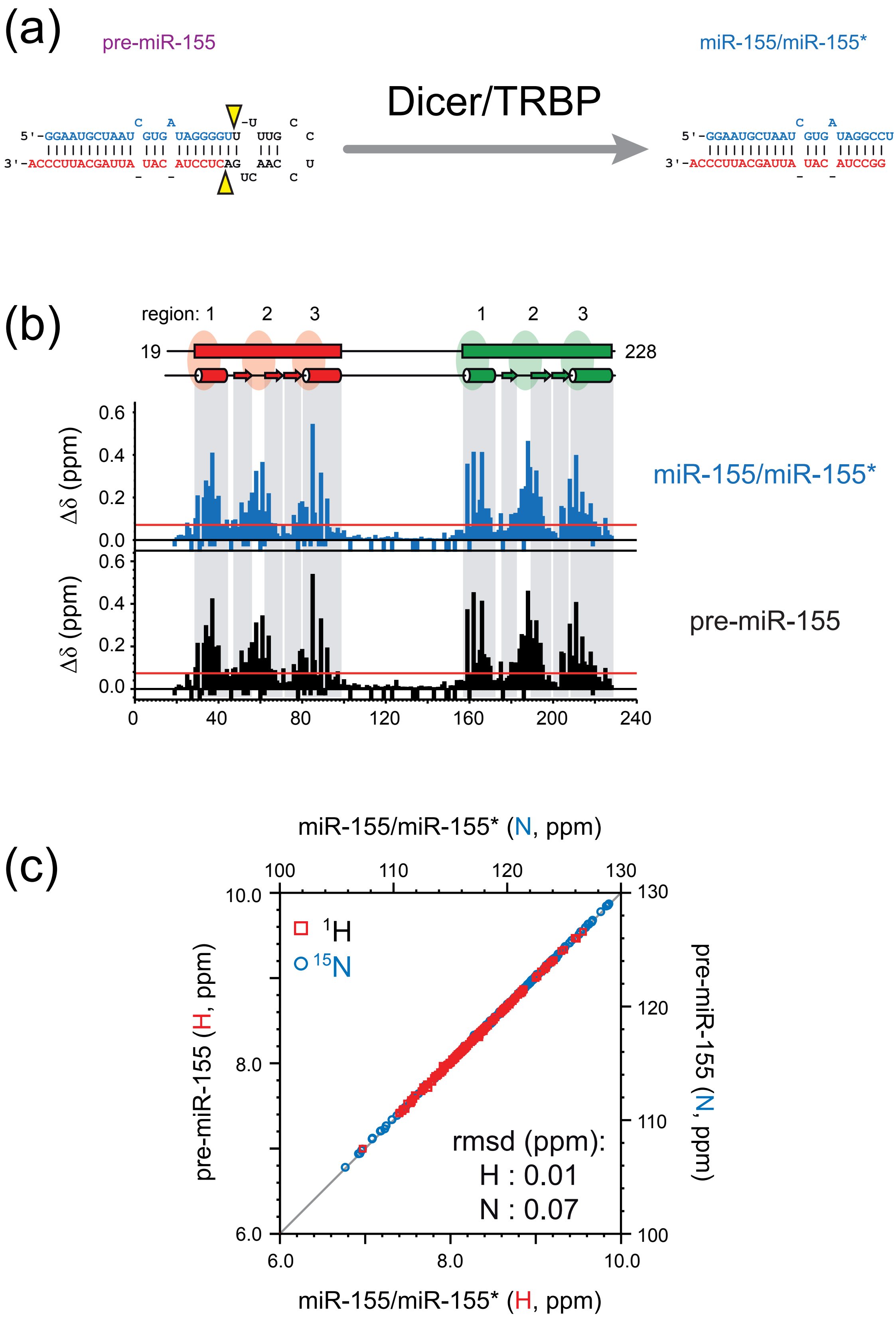
Supplementary Figure S3



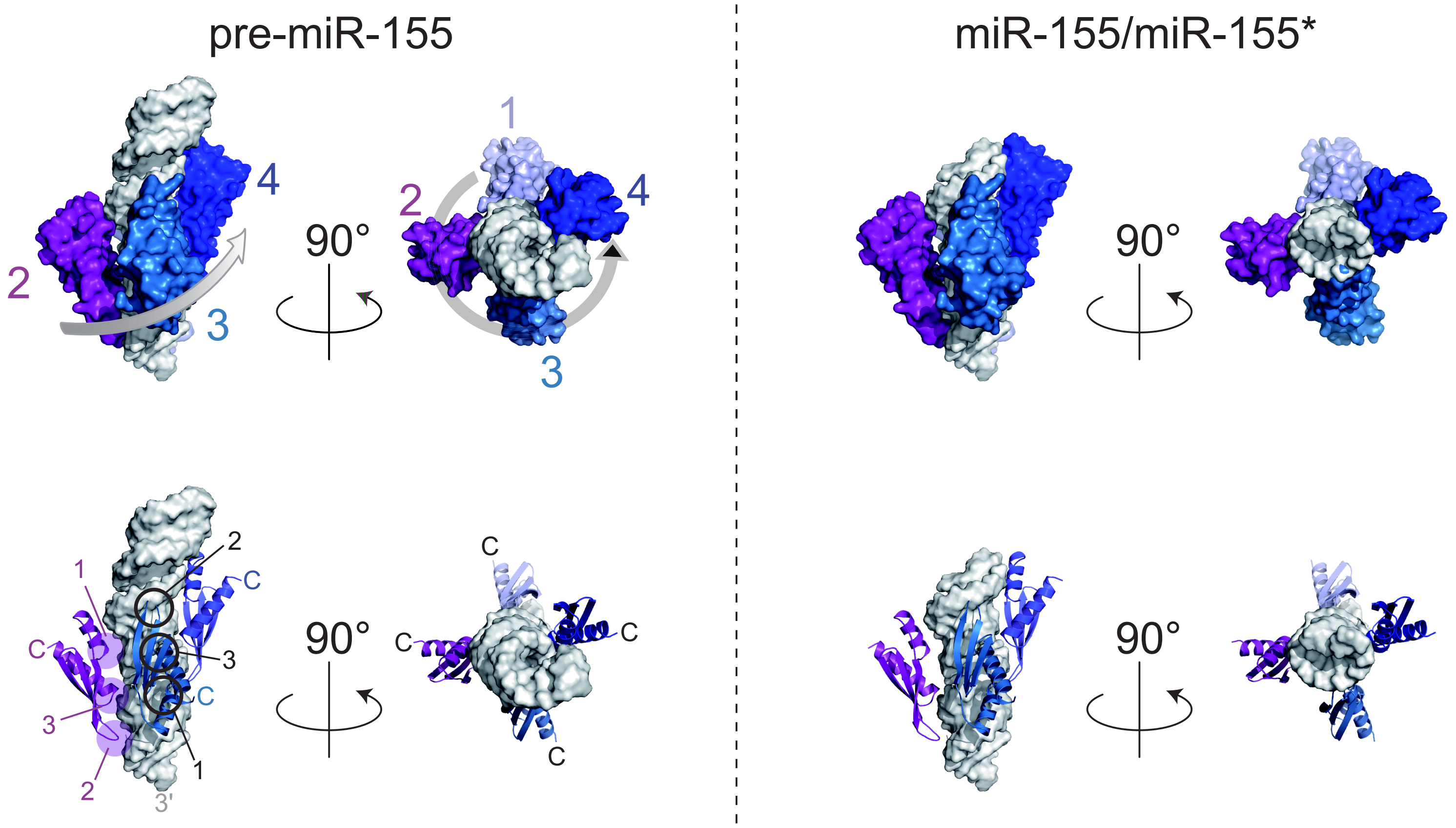
Supplementary Figure S4



Supplementary Figure S5



Supplementary Figure S6



Supplementary Table 1 – Summary of svAUC data

	Experimental values					Calculated values							
Sample	s_{exp} **(S)**	M_{b} **(kDa)**	$\frac{A_{280}}{A_{260}}$	$\frac{\Delta J}{A_{280}}$	$\frac{\Delta J}{A_{260}}$	**Protein:RNA**	M_{calc} **(kDa)**	$\bar{v}$ **(mL/g)**	s_{calc} **(S)**	$M_{\mathrm{b,calc}}$ **(kDa)**	$\frac{A_{280}}{A_{260}}$	$\frac{\Delta J}{A_{280}}$	$\frac{\Delta J}{A_{260}}$
TRBP-D12	2.1 (0.1)	5.7	1.73	5.7	9.9	1:0	22.3	0.74	2.1	5.7			
pre-miR-155	4.4 (0.2)	9.1	0.46	0.3	0.14	0:1	19.4	0.508	4.3	9.2			
Complex	5.9 (0.1)	18.3	0.49	0.8	0.41	1:1	41.9	0.63	4.9 - 5.5	14.9	0.47	0.53	0.25
						2:1	61.5	0.67	5.7 - 6.4	20.6	0.49	0.75	0.37
						1:2	64.2	0.59	7.0 - 7.8	24.0	0.47	0.42	0.20

s_{exp}, experimental sedimentation coefficient with standard deviation in parentheses; M_{b}, experimental buoyant mass determined with the non-interacting species analysis; A_{280} and A_{260}, absorbance at 280 nm and 260 nm, respectively; ΔJ, interference fringe shift; M_{calc}, $\bar{v}$ and $M_{\mathrm{b,calc}}$, mass, partial specific volume and buoyant mass predicted from molecular composition; lower and upper s_{calc} limits for the TRBP-D12/pre-miR-155 complex were calculated using frictional coefficients of 1.4 (values calculated for free TRBP-D12 and free pre-miR-155) or 1.25 (corresponding to globular shape), respectively. The ratios for A_{280}, A_{260} and ΔJ are calculated using the experimentally determined values for TRBP-D12 and pre-miR-155.

Supplementary Table 2 – Summary of ITC results

	Syringe	**Cell**	**K_d (μM)**		***N***	**ΔH (kcal mol^{-1})**		**ΔS (cal mol^{-1} K^{-1})**	
1	pre-miR-155	TRBP-D1		15.5 (0.8)	3.9 (0.2)		-6.0 (0.2)		1.9 (0.2)
2	pre-miR-155	TRBP-D2		3.7 (0.5)	4.1 (0.1)		-9.4 (0.2)		-6.8 (0.2)
3	pre-miR-155	TRBP-D12		3.4 (0.5)	2.0 (0.1)		-16.1 (0.3)		-28.8 (0.3)
4	TRBP-D12	pre-miR-155	1	0.5 (0.03)	n/a	1	-17.7 (0.1)	1	-30.5 (0.2)
			2	6.2 (0.2)	n/a	2	-8.3 (0.1)	2	-4.2 (0.3)
5	TRBP-D12	miR-155/miR155*	1	1.0 (0.7)	n/a	1	-20. 6 (0.2)	1	-41.9 (0.1)
			2	5.7 (0.3)	n/a	2	-2.3 (0.2)	2	16.3 (0.6)

The values given correspond to the combined average (and standard deviation) of the fitting parameters from a minimum of two experiments. **ROWS 1-3**: The binding isotherms resulting from the titration of pre-miR-155 into different TRBP constructs were evaluated using a single site model where *N* refers to the stoichiometry. The standard deviations associated with the reported K_d, ΔH and ΔS values were derived from experimental repetition ($n \geq 2$). **ROWS 4 and 5**: The binding isotherm resulting from the titration of TRBP-D12 into different RNA samples was evaluated using a 2-site sequential binding model. $K_{d,i}$, ΔH_i and ΔS_i values are given for each binding event (i = 1 or 2). Each experiment was repeated twice. The standard deviation for each fitted parameter is indicated in parenthesis and was determined by running 100 Monte Carlo simulations of each pair of ITC curves followed by a global fitting of each pair of experiments.

20.4 Expressions of total concentration in ITC titrations

This part explicit the calculation of the total concentration in the cell during an ITC experiment. The following notation are used here for the demonstration of the different concentration expressions found in the literature

Total concentration of macromolecules under the continuous mixing hypothesis By conservation of the mass during the injection i in the control volume we have :

$$\Delta n_{\text{titrant}} = n_{\text{titrant coming}} - n_{\text{titrant leaving}} \tag{20.1}$$

so, using the concentrations :

$$(c + dc)V - cV = c_0 dv - cd \tag{20.2}$$

which leads to a linear differential equation of the first degree :

$$\frac{dc}{dv} + \frac{1}{V}c - \frac{1}{V}c_0 = 0 \tag{20.3}$$

So, for any injection i :

$$C_{\text{titrant}}(i) = C_{\text{titrant in syringe}} \left(1 - exp\left(-\frac{v(i)}{V}\right)\right) \tag{20.4}$$

Similarly, the related expression for the analyte is :

$$C_{\text{analyte}}(i) = C_{\text{analyte in cell}} \left(exp\left(-\frac{v(i)}{V}\right)\right) \tag{20.5}$$

Total concentration of macromolecules under the no mixing hypothesis It is considered here that at injection i the expelled volume is only composed by the solution present in the cell before this injection. For an injected volume v_i during the titration i satisfying $v_i \ll V$, using the same notation as the previous paragraph, we have

$$\Delta n_{\text{titrant}} = n_{\text{titrant coming}} - n_{\text{titrant leaving}} \tag{20.6}$$

so, using the concentrations :

$$c_{i+1}V - c_iV = c_0V - c_iv_i \tag{20.7}$$

Table 20.1: Notations

Notation	Definition
I	number of injections
i	injection number in $[\![1, I]\!]$
$v(i)$	the injection volume of the injection i
dv	an elementary injection volume of titrant
c	the concentration of titrant in the cell before injection
$n_{titrant}$	the molar amount of titrant in the control volume
V	volume of the control volume $V = V_{\text{cell}} - V_{\text{paddle}}$

which can be written as a recurrent relation :

$$\frac{c_{i+1}}{c_i} = \left(1 - \frac{v_i}{V}\right) \tag{20.8}$$

and by recurrence :

$$C_{\text{titrant}}(i) = C_{\text{titrant in syringe}} \left(1 - \left(1 - \frac{v(i)}{V}\right)^i\right) \tag{20.9}$$

Similarly, the related expression for the analyte is :

$$C_{\text{analyte}}(i) = C_{\text{analyte in cell}} \left(1 - \frac{v(i)}{V}\right)^i \tag{20.10}$$

Another approach MicroCal uses another approach, based on an estimation on the concentration of the expelled volume as the mean of the concentration of the cell before the first injection and after the injection :

$$c_i = c_0 V + \frac{1}{2}(c_i + c_0) v_i \tag{20.11}$$

so :

$$C_{\text{analyte}}(i) = C_0 \left(\frac{1 - \frac{v(i)}{2V}}{1 + \frac{v(i)}{2V}}\right) \tag{20.12}$$

Bibliography

[1] *Hardware Manual for the DAWN EOS Light Scattering Instrument*. Wyatt Technology, 2002.

[2] *Technical Resource Guide. Fluorescence Polarization (Fourth edition)*. Invitrogen, 2006.

[3] *ASTRA V User's Guide*. Wyatt Technology, 2008.

[4] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, Dec 1974.

[5] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *J. Mol. Biol.*, 215(3):403–410, Oct 1990.

[6] V. Ambros, B. Bartel, D. P. Bartel, C. B. Burge, J. C. Carrington, X. Chen, G. Dreyfuss, S. R. Eddy, S. Griffiths-Jones, M. Marshall, M. Matzke, G. Ruvkun, and T. Tuschl. A uniform system for microRNA annotation. *RNA*, 9(3):277–279, Mar 2003.

[7] Y. Ando, Y. Tomaru, A. Morinaga, A. M. Burroughs, H. Kawaji, A. Kubosaki, R. Kimura, M. Tagata, Y. Ino, H. Hirano, J. Chiba, H. Suzuki, P. Carninci, and Y. Hayashizaki. Nuclear pore complex protein mediated nuclear localization of dicer protein in human cells. *PLoS ONE*, 6(8):e23385, 2011.

[8] No authors listed. Recommendations for measurement and presentation of biochemical equilibrium data. *Q. Rev. Biophys.*, 9(3):439–456, Aug 1976.

[9] J. E. Babiarz, J. G. Ruby, Y. Wang, D. P. Bartel, and R. Blelloch. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev.*, 22(20):2773–2785, Oct 2008.

[10] A. Balbo, K. H. Minor, C. A. Velikovsky, R. A. Mariuzza, C. B. Peterson, and P. Schuck. Studying multiprotein complexes by multisignal sedimentation velocity analytical ultracentrifugation. *Proc. Natl. Acad. Sci. U.S.A.*, 102(1):81–86, Jan 2005.

[11] V. Ball and C. Maechling. Isothermal microcalorimetry to investigate non specific interactions in biophysical chemistry. *Int J Mol Sci*, 10(8):3283–3315, Oct 2009.

[12] P. Barraud and F. H. Allain. ADAR proteins: double-stranded RNA and Z-DNA binding domains. *Curr. Top. Microbiol. Immunol.*, 353:35–60, 2012.

[13] P. Barraud, S. Emmerth, Y. Shimada, H. R. Hotz, F. H. Allain, and M. Buhler. An extended dsRBD with a novel zinc-binding motif mediates nuclear retention of fission yeast Dicer. *EMBO J.*, 30(20):4223–4235, Oct 2011.

[14] P. Barraud, B. S. Heale, M. A. O'Connell, and F. H. Allain. Solution structure of the N-terminal dsRBD of Drosophila ADAR and interaction studies with RNA. *Biochimie*, 94(7):1499–1509, Jul 2012.

[15] M. P. Benoit, L. Imbert, A. Palencia, J. Perard, C. Ebel, J. Boisbouvier, and M. J. Plevin. The RNA-binding region of human TRBP interacts with microRNA precursors through two independent domains. *Nucleic Acids Res.*, Feb 2013.

[16] M. P. Benoit and M. J. Plevin. Backbone resonance assignments of the micro-RNA precursor binding region of human TRBP. *Biomol NMR Assign*, Aug 2012.

[17] E. Berezikov, W. J. Chung, J. Willis, E. Cuppen, and E. C. Lai. Mammalian mirtron genes. *Mol. Cell*, 28(2):328–336, Oct 2007.

[18] M. L. Bianconi. Calorimetry of enzyme-catalyzed reactions. *Biophys. Chem.*, 126(1-3):59–64, Mar 2007.

[19] T. Blevins, R. Rajeswaran, P. V. Shivaprasad, D. Beknazariants, A. Si-Ammour, H. S. Park, F. Vazquez, D. Robertson, F. Meins, T. Hohn, and M. M. Pooggin. Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Res.*, 34(21):6233–6246, 2006.

[20] G. F. Bonifacio, T. Brown, G. L. Conn, and A. N. Lane. Comparison of the electrophoretic and hydrodynamic properties of DNA and RNA oligonucleotide duplexes. *Biophys. J.*, 73(3):1532–1538, Sep 1997.

[21] J. Brennecke, A. Stark, R. B. Russell, and S. M. Cohen. Principles of microRNA-target recognition. *PLoS Biol.*, 3(3):e85, Mar 2005.

[22] R. Brooks, C. R. Eckmann, and M. F. Jantsch. The double-stranded RNA-binding domains of Xenopus laevis ADAR1 exhibit different RNA-binding behaviors. *FEBS Lett.*, 434(1-2):121–126, Aug 1998.

[23] P. Burdisso, I. P. Suarez, N. G. Bologna, J. F. Palatnik, B. Bersch, and R. M. Rasia. Second double-stranded RNA binding domain of dicer-like ribonuclease 1: structural and biochemical characterization. *Biochemistry*, 51(51):10159–10166, Dec 2012.

[24] K. P. Burnham and D. R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed.)*. Springer-Verlag, 2002.

[25] D. Burnouf, E. Ennifar, S. Guedich, B. Puffer, G. Hoffmann, G. Bec, F. Disdier, M. Baltzinger, and P. Dumas. kinITC: a new method for obtaining joint thermodynamic and kinetic data by isothermal titration calorimetry. *J. Am. Chem. Soc.*, 134(1):559–565, Jan 2012.

[26] L. L. Burns-Hamuro, P. M. Dalessio, and I. J. Ropson. Replacement of proline with valine does not remove an apparent proline isomerization-dependent folding event in CRABP I. *Protein Sci.*, 13(6):1670–1676, Jun 2004.

[27] N. Bushati and S. M. Cohen. microRNA functions. *Annu. Rev. Cell Dev. Biol.*, 23:175–205, 2007.

[28] D. Castanotto, K. Sakurai, R. Lingeman, H. Li, L. Shively, L. Aagaard, H. Soifer, A. Gatignol, A. Riggs, and J. J. Rossi. Combinatorial delivery of small interfering RNAs reduces RNAi efficacy by selective incorporation into RISC. *Nucleic Acids Res.*, 35(15):5154–5164, 2007.

[29] J Cavanagh, W. J. Fairbrother, A. G. Palmer III, and Rance M. Skelton, N. J. *Protein NMR Spectroscopy, Second Edition : Principles and Practice* . Elsevier, 2007.

[30] E. S. Cenik, R. Fukunaga, G. Lu, R. Dutcher, Y. Wang, T. M. Tanaka Hall, and P. D. Zamore. Phosphate and R2D2 restrict the substrate specificity of Dicer-2, an ATP-driven ribonuclease. *Mol. Cell*, 42(2):172–184, Apr 2011.

[31] S. Chakravarthy, S. H. Sternberg, C. A. Kellenberger, and J. A. Doudna. Substrate-specific kinetics of Dicer-catalyzed RNA processing. *J. Mol. Biol.*, 404(3):392–402, Dec 2010.

[32] Chamberlin. THE METHOD OF MULTIPLE WORKING HYPOTHESES. *Science*, 15(366):92–96, Feb 1890.

[33] T. C. Chamberlin. The Method of Multiple Working Hypotheses: With this method the dangers of parental affection for a favorite theory can be circumvented. *Science*, 148(3671):754–759, May 1965.

[34] K. Y. Chang and A. Ramos. The double-stranded RNA-binding motif, a versatile macromolecular docking platform. *FEBS J.*, 272(9):2109–2117, May 2005.

[35] T. P. Chendrimada, R. I. Gregory, E. Kumaraswamy, J. Norman, N. Cooch, K. Nishikura, and R. Shiekhattar. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, 436(7051):740–744, Aug 2005.

[36] S. Chiliveri and M.V. Deshmukh. poster presentation ENC. 2013.

[37] W. J. Chung, P. Agius, J. O. Westholm, M. Chen, K. Okamura, N. Robine, C. S. Leslie, and E. C. Lai. Computational and experimental identification of mirtrons in Drosophila melanogaster and Caenorhabditis elegans. *Genome Res.*, 21(2):286–300, Feb 2011.

[38] D. Cifuentes, H. Xue, D. W. Taylor, H. Patnode, Y. Mishima, S. Cheloufi, E. Ma, S. Mane, G. J. Hannon, N. D. Lawson, S. A. Wolfe, and A. J. Giraldez. A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science*, 328(5986):1694–1698, Jun 2010.

[39] J. L. Cole, J. J. Correia, and W. F. Stafford. The use of analytical sedimentation velocity to extract thermodynamic linkage. *Biophys. Chem.*, 159(1):120–128, Nov 2011.

[40] A. Cook and E. Conti. Dicer measures up. *Nat. Struct. Mol. Biol.*, 13(3):190–192, Mar 2006.

[41] A. Cooper and C. M. Johnson. Introduction to microcalorimetry and biomolecular energetics. *Methods Mol. Biol.*, 22:109–124, 1994.

[42] G. P. Cosentino, S. Venkatesan, F. C. Serluca, S. R. Green, M. B. Mathews, and N. Sonenberg. Double-stranded-RNA-dependent protein kinase and TAR RNA-binding protein form homo- and heterodimers in vivo. *Proc. Natl. Acad. Sci. U.S.A.*, 92(21):9445–9449, Oct 1995.

[43] B. Czech, C. D. Malone, R. Zhou, A. Stark, C. Schlingeheyde, M. Dus, N. Perrimon, M. Kellis, J. A. Wohlschlegel, R. Sachidanandam, G. J. Hannon, and J. Brennecke. An endogenous small interfering RNA pathway in Drosophila. *Nature*, 453(7196):798–802, Jun 2008.

[44] A. Daher, G. Laraki, M. Singh, C. E. Melendez-Pena, S. Bannwarth, A. H. Peters, E. F. Meurs, R. E. Braun, R. C. Patel, and A. Gatignol. TRBP control of PACT-induced phosphorylation of protein kinase R is reversed by stress. *Mol. Cell. Biol.*, 29(1):254–265, Jan 2009.

[45] A. Daher, M. Longuet, D. Dorin, F. Bois, E. Segeral, S. Bannwarth, P. L. Battisti, D. F. Purcell, R. Benarous, C. Vaquero, E. F. Meurs, and A. Gatignol. Two dimerization domains in the trans-activation response RNA-binding protein (TRBP) individually reverse the protein kinase R inhibition of HIV-1 long terminal repeat expression. *J. Biol. Chem.*, 276(36):33899–33905, Sep 2001.

[46] J. Dam and P. Schuck. Sedimentation velocity analysis of heterogeneous protein-protein interactions: sedimentation coefficient distributions c(s) and asymptotic boundary profiles from Gilbert-Jenkins theory. *Biophys. J.*, 89(1):651–666, Jul 2005.

[47] S. M. Daniels and A. Gatignol. The multiple functions of TRBP, at the hub of cell responses to viruses, stress, and cancer. *Microbiol. Mol. Biol. Rev.*, 76(3):652–666, Sep 2012.

[48] S. M. Daniels, C. E. Melendez-Pena, R. J. Scarborough, A. Daher, H. S. Christensen, M. El Far, D. F. Purcell, S. Laine, and A. Gatignol. Characterization of the TRBP domain required for dicer interaction and function in RNA interference. *BMC Mol. Biol.*, 10:38, 2009.

[49] R. Das, J. Karanicolas, and D. Baker. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat. Methods*, 7(4):291–294, Apr 2010.

[50] E. de Castro, C. J. Sigrist, A. Gattiker, V. Bulliard, P. S. Langendijk-Genevaux, E. Gasteiger, A. Bairoch, and N. Hulo. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.*, 34(Web Server issue):W362–365, Jul 2006.

[51] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, 6:277–293, Nov 1995.

[52] A. Deleris, J. Gallego-Bartolome, J. Bao, K. D. Kasschau, J. C. Carrington, and O. Voinnet. Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science*, 313(5783):68–71, Jul 2006.

[53] K. Denbigh. *The Principes of Chemical Equilibriums*. Cambridge University Press, Fourth edition edition, 1997.

[54] A. M. Denli, B. B. Tops, R. H. Plasterk, R. F. Ketting, and G. J. Hannon. Processing of primary microRNAs by the Microprocessor complex. *Nature*, 432(7014):231–235, Nov 2004.

[55] M. Dlaki. DUF283 domain of Dicer proteins has a double-stranded RNA-binding fold. *Bioinformatics*, 22(22):2711–2714, Nov 2006.

[56] N. Doi, S. Zenno, R. Ueda, H. Ohki-Hamazaki, K. Ui-Tei, and K. Saigo. Short-interfering-RNA-mediated gene silencing in mammalian cells requires Dicer and eIF2C translation initiation factors. *Curr. Biol.*, 13(1):41–46, Jan 2003.

[57] Z. Dong, M. H. Han, and N. Fedoroff. The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. *Proc. Natl. Acad. Sci. U.S.A.*, 105(29):9970–9975, Jul 2008.

[58] M. Doyle and M. F. Jantsch. New and old roles of the double-stranded RNA-binding domain. *J. Struct. Biol.*, 140(1-3):147–153, 2002.

[59] T. Du and P. D. Zamore. microPrimer: the biogenesis and function of microRNA. *Development*, 132(21):4645–4652, Nov 2005.

[60] Z. Du, J. K. Lee, R. Tjhen, R. M. Stroud, and T. L. James. Structural and biochemical insights into the dicing mechanism of mouse Dicer: a conserved lysine is critical for dsRNA cleavage. *Proc. Natl. Acad. Sci. U.S.A.*, 105(7):2391–2396, Feb 2008.

[61] M. S. Duxbury, S. W. Ashley, and E. E. Whang. RNA interference: a mammalian SID-1 homologue enhances siRNA uptake and gene silencing efficacy in human cells. *Biochem. Biophys. Res. Commun.*, 331(2):459–463, Jun 2005.

[62] A. L. Eamens, K. W. Kim, S. J. Curtin, and P. M. Waterhouse. DRB2 is required for microRNA biogenesis in Arabidopsis thaliana. *PLoS ONE*, 7(4):e35933, 2012.

[63] A. L. Eamens, N. A. Smith, S. J. Curtin, M. B. Wang, and P. M. Waterhouse. The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *RNA*, 15(12):2219–2235, Dec 2009.

[64] A. L. Eamens, K. Wook Kim, and P. M. Waterhouse. DRB2, DRB3 and DRB5 function in a non-canonical microRNA pathway in Arabidopsis thaliana. *Plant Signal Behav*, 7(10):1224–1229, Oct 2012.

[65] R. C. Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, 32(5):1792–1797, 2004.

[66] P.H.C. Eilers and Boelens H.F.M. Baseline Correction with Asymmetric Least Squares Smoothing. Oct 2005.

[67] H. Eisenberg. Modern analytical ultracentrifugation in protein science: look forward, not back. *Protein Sci.*, 12(11):2647–2649, Nov 2003.

[68] M. O. Elhassan, J. Christie, and M. S. Duxbury. Homo sapiens systemic RNA interference-defective-1 transmembrane family member 1 (SIDT1) protein mediates contact-dependent small RNA transfer and microRNA-21-driven chemoresistance. *J. Biol. Chem.*, 287(8):5267–5277, Feb 2012.

[69] L.P. Elliott and Brook B.W. Revisiting Chamberlin: Multiple Working Hypotheses for the 21st Century. *BioScience*, 57(7):608–614, Jul 2007.

[70] T. S. Elton, H. Selemon, S. M. Elton, and N. L. Parinandi. Regulation of the MIR155 host gene in physiological and pathological processes. *Gene*, Dec 2012.

[71] C. Ender, A. Krek, M. R. Friedlander, M. Beitzinger, L. Weinmann, W. Chen, S. Pfeffer, N. Rajewsky, and G. Meister. A human snoRNA with microRNA-like functions. *Mol. Cell*, 32(4):519–528, Nov 2008.

[72] M. Faller, D. Toso, M. Matsunaga, I. Atanasov, R. Senturia, Y. Chen, Z. H. Zhou, and F. Guo. DGCR8 recognizes primary transcripts of microRNAs through highly cooperative binding and formation of higher-order structures. *RNA*, 16(8):1570–1583, Aug 2010.

[73] Y. Fang and D. L. Spector. Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants. *Curr. Biol.*, 17(9):818–823, May 2007.

[74] N. A. Farrow, R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay, and L. E. Kay. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry*, 33(19):5984–6003, May 1994.

[75] N. A. Farrow, O. Zhang, J. D. Forman-Kay, and L. E. Kay. A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR*, 4(5):727–734, Sep 1994.

[76] N. A. Farrow, O. Zhang, J. D. Forman-Kay, and L. E. Kay. Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer. *Biochemistry*, 34(3):868–878, Jan 1995.

[77] A. Favier and B. Brutscher. Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J. Biomol. NMR*, 49(1):9–15, Jan 2011.

[78] A. L. Feig. Studying RNA-RNA and RNA-protein interactions by isothermal titration calorimetry. *Meth. Enzymol.*, 468:409–422, 2009.

[79] S. L. Fernandez-Valverde, R. J. Taft, and J. S. Mattick. Dynamic isomiR regulation in Drosophila development. *RNA*, 16(10):1881–1888, Oct 2010.

[80] I. Fierro-Monti and M. B. Mathews. Proteins binding to duplexed RNA: one motif, multiple functions. *Trends Biochem. Sci.*, 25(5):241–246, May 2000.

[81] W. Filipowicz, S. N. Bhattacharyya, and N. Sonenberg. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.*, 9(2):102–114, Feb 2008.

[82] S. E. Fischer. Small RNA-mediated gene silencing pathways in C. elegans. *Int. J. Biochem. Cell Biol.*, 42(8):1306–1315, Aug 2010.

[83] K. Forstemann, M. D. Horwich, L. Wee, Y. Tomari, and P. D. Zamore. Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell*, 130(2):287–297, Jul 2007.

[84] K. Forstemann, Y. Tomari, T. Du, V. V. Vagin, A. M. Denli, D. P. Bratu, C. Klattenhoff, W. E. Theurkauf, and P. D. Zamore. Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biol.*, 3(7):e236, Jul 2005.

[85] W. D. Foulkes, A. Bahubeshi, N. Hamel, B. Pasini, S. Asioli, G. Baynam, C. S. Choong, A. Charles, R. P. Frieder, M. K. Dishop, N. Graf, M. Ekim, D. Bouron-Dal Soglio, J. Arseneau, R. H. Young, N. Sabbaghian, A. Srivastava, M. D. Tischkowitz, and J. R. Priest. Extending the phenotypes associated with DICER1 mutations. *Hum. Mutat.*, 32(12):1381–1384, Dec 2011.

[86] E. Freire, A. Schon, and A. Velazquez-Campoy. Isothermal titration calorimetry: general formalism using binding polynomials. *Meth. Enzymol.*, 455:127–155, 2009.

[87] M. W. Freyer and E. A. Lewis. Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol.*, 84:79–113, 2008.

[88] R. C. Friedman, K. K. Farh, C. B. Burge, and D. P. Bartel. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, 19(1):92–105, Jan 2009.

[89] D. Frishman and P. Argos. Knowledge-based protein secondary structure assignment. *Proteins*, 23(4):566–579, Dec 1995.

[90] J. Fritz, A. Strehblow, A. Taschner, S. Schopoff, P. Pasierbek, and M. F. Jantsch. RNA-regulated interaction of transportin-1 and exportin-5 with the double-stranded RNA-binding domain regulates nucleocytoplasmic shuttling of ADAR1. *Mol. Cell. Biol.*, 29(6):1487–1497, Mar 2009.

[91] R. Fukunaga, B. W. Han, J. H. Hung, J. Xu, Z. Weng, and P. D. Zamore. Dicer partner proteins tune the length of mature miRNAs in flies and mammals. *Cell*, 151(3):533–546, Oct 2012.

[92] R. Fukunaga and P Zamore. *Loquacious, a Dicer Partner Protein, Functions in Both the MicroRNA and siRNA Pathways*, volume 32 of *The Enzymes, Chapter 2*. Elsevier.

[93] B. Furtig, C. Richter, J. Wohnert, and H. Schwalbe. NMR spectroscopy of RNA. *Chembiochem*, 4(10):936–962, Oct 2003.

[94] R. Garzon and C. M. Croce. MicroRNAs and cancer: introduction. *Semin. Oncol.*, 38(6):721–723, Dec 2011.

[95] V. Gasciolli, A. C. Mallory, D. P. Bartel, and H. Vaucheret. Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr. Biol.*, 15(16):1494–1500, Aug 2005.

[96] A. Gatignol, C. Buckler, and K. T. Jeang. Relatedness of an RNA-binding motif in human immunodeficiency virus type 1 TAR RNA-binding protein TRBP to human P1/dsI kinase and Drosophila staufen. *Mol. Cell. Biol.*, 13(4):2193–2202, Apr 1993.

[97] A. Gatignol, A. Buckler-White, B. Berkhout, and K. T. Jeang. Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. *Science*, 251(5001):1597–1600, Mar 1991.

[98] M. Gerstein and R. B. Altman. Average core structures and variability measures for protein families: application to the immunoglobulins. *J. Mol. Biol.*, 251(1):161–175, Aug 1995.

[99] M. Ghildiyal and P. D. Zamore. Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.*, 10(2):94–108, Feb 2009.

[100] M. Gironella, M. Seux, M. J. Xie, C. Cano, R. Tomasini, J. Gommeaux, S. Garcia, J. Nowak, M. L. Yeung, K. T. Jeang, A. Chaix, L. Fazli, Y. Motoo, Q. Wang, P. Rocchi, A. Russo, M. Gleave, J. C. Dagorn, J. L. Iovanna, A. Carrier, M. J. Pebusque, and N. J. Dusetti. Tumor protein 53-induced nuclear protein 1 expression is repressed by miR-155, and its restoration inhibits pancreatic tumor development. *Proc. Natl. Acad. Sci. U.S.A.*, 104(41):16170–16175, Oct 2007.

[101] G. Glatting, P. Kletting, S. N. Reske, K. Hohl, and C. Ring. Choosing the optimal fit function: comparison of the Akaike information criterion and the F-test. *Med Phys*, 34(11):4285–4292, Nov 2007.

[102] N. Goudreau, F. Cornille, M. Duchesne, F. Parker, B. Tocque, C. Garbay, and B. P. Roques. NMR structure of the N-terminal SH3 domain of GRB2 and its complex with a proline-rich peptide from Sos. *Nat. Struct. Biol.*, 1(12):898–907, Dec 1994.

[103] B. J. Grant, A. P. Rodrigues, K. M. ElSawy, J. A. McCammon, and L. S. Caves. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*, 22(21):2695–2696, Nov 2006.

[104] S. Graslund, P. Nordlund, J. Weigelt, B. M. Hallberg, J. Bray, O. Gileadi, S. Knapp, U. Oppermann, C. Arrowsmith, R. Hui, J. Ming, S. dhe Paganon, H. W. Park, A. Savchenko, A. Yee, A. Edwards, R. Vincentelli, C. Cambillau, R. Kim, S. H. Kim, Z. Rao, Y. Shi, T. C. Terwilliger, C. Y. Kim, L. W. Hung, G. S. Waldo, Y. Peleg,

S. Albeck, T. Unger, O. Dym, J. Prilusky, J. L. Sussman, R. C. Stevens, S. A. Lesley, I. A. Wilson, A. Joachimiak, F. Collart, I. Dementieva, M. I. Donnelly, W. H. Eschenfeldt, Y. Kim, L. Stols, R. Wu, M. Zhou, S. K. Burley, J. S. Emtage, J. M. Sauder, D. Thompson, K. Bain, J. Luz, T. Gheyi, F. Zhang, S. Atwell, S. C. Almo, J. B. Bonanno, A. Fiser, S. Swaminathan, F. W. Studier, M. R. Chance, A. Sali, T. B. Acton, R. Xiao, L. Zhao, L. C. Ma, J. F. Hunt, L. Tong, K. Cunningham, M. Inouye, S. Anderson, H. Janjua, R. Shastry, C. K. Ho, D. Wang, H. Wang, M. Jiang, G. T. Montelione, D. I. Stuart, R. J. Owens, S. Daenke, A. Schutz, U. Heinemann, S. Yokoyama, K. Bussow, and K. C. Gunsalus. Protein production and purification. *Nat. Methods*, 5(2):135–146, Feb 2008.

[105] J. A. Gredell, M. J. Dittmer, M. Wu, C. Chan, and S. P. Walton. Recognition of siRNA asymmetry by TAR RNA binding protein. *Biochemistry*, 49(14):3148–3155, Apr 2010.

[106] R. I. Gregory, T. P. Chendrimada, N. Cooch, and R. Shiekhattar. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell*, 123(4):631–640, Nov 2005.

[107] J. P. Grolier and J. M. del Rio. On the physical meaning of the isothermal titration calorimetry measurements in calorimeters with full cells. *Int J Mol Sci*, 10(12):5296–5325, Dec 2009.

[108] V. Gupta, X. Huang, and R. C. Patel. The carboxy-terminal, M3 motifs of PACT and TRBP have opposite effects on PKR activity. *Virology*, 315(2):283–291, Oct 2003.

[109] A. D. Haase, L. Jaskiewicz, H. Zhang, S. Laine, R. Sack, A. Gatignol, and W. Filipowicz. TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep.*, 6(10):961–967, Oct 2005.

[110] B.D. Hames. *Gel Electrophoresis of Proteins: A Practical Approach*. The Practical Approach Series. Oxford University Press, Incorporated, 1998.

[111] J. Han, Y. Lee, K. H. Yeom, Y. K. Kim, H. Jin, and V. N. Kim. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.*, 18(24):3016–3027, Dec 2004.

[112] J. Han, Y. Lee, K. H. Yeom, J. W. Nam, I. Heo, J. K. Rhee, S. Y. Sohn, Y. Cho, B. T. Zhang, and V. N. Kim. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, 125(5):887–901, Jun 2006.

[113] M. H. Han, S. Goud, L. Song, and N. Fedoroff. The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc. Natl. Acad. Sci. U.S.A.*, 101(4):1093–1098, Jan 2004.

[114] Y. H. Han, Y. J. Luo, Q. Wu, J. Jovel, X. H. Wang, R. Aliyari, C. Han, W. X. Li, and S. W. Ding. RNA-based immunity terminates viral infection in adult Drosophila in the absence of viral suppression of RNA interference: characterization of viral small interfering RNA populations in wild-type and mutant flies. *J. Virol.*, 85(24):13153–13163, Dec 2011.

[115] D.B. Hand. The refractivity of protein solutions. *J. Biol. Chem.*, 108:703–707, 1935.

[116] M. M. Hanna, L. Bentsen, M. Lucido, and A. Sapre. RNA-protein crosslinking with photoreactive nucleotide analogs. *Methods Mol. Biol.*, 118:21–33, 1999.

[117] D. F. Hansen and J. J. Led. Implications of using approximate Bloch-McConnell equations in NMR analyses of chemically exchanging systems: application to the electron self-exchange of plastocyanin. *J. Magn. Reson.*, 163(2):215–227, Aug 2003.

[118] S. E. Harding and P. Johnson. The concentration-dependence of macromolecular parameters. *Biochem. J.*, 231(3):543–547, Nov 1985.

[119] J. V. Hartig, S. Esslinger, R. Bottcher, K. Saito, and K. Forstemann. Endo-siRNAs depend on a new isoform of loquacious and target artificially introduced, high-copy sequences. *EMBO J.*, 28(19):2932–2944, Oct 2009.

[120] J. V. Hartig and K. Forstemann. Loqs-PD and R2D2 define independent pathways for RISC generation in Drosophila. *Nucleic Acids Res.*, 39(9):3836–3851, May 2011.

[121] M. A. Havens, A. A. Reich, D. M. Duelli, and M. L. Hastings. Biogenesis of mammalian microRNAs by a non-canonical processing pathway. *Nucleic Acids Res.*, 40(10):4626–4640, May 2012.

[122] A. Hiraguri, R. Itoh, N. Kondo, Y. Nomura, D. Aizawa, Y. Murai, H. Koiwa, M. Seki, K. Shinozaki, and T. Fukuhara. Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in Arabidopsis thaliana. *Plant Mol. Biol.*, 57(2):173–188, Jan 2005.

[123] E. Hitti, A. Neunteufl, and M. F. Jantsch. The double-stranded RNA-binding protein X1rbpa promotes RNA strand annealing. *Nucleic Acids Res.*, 26(19):4382–4388, Oct 1998.

[124] L. Holm and P. Rosenstrom. Dali server: conservation mapping in 3D. *Nucleic Acids Res.*, 38(Web Server issue):W545–549, Jul 2010.

[125] M. D. Horwich, C. Li, C. Matranga, V. Vagin, G. Farley, P. Wang, and P. D. Zamore. The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.*, 17(14):1265–1272, Jul 2007.

[126] J. C. Houtman, P. H. Brown, B. Bowden, H. Yamaguchi, E. Appella, L. E. Samelson, and P. Schuck. Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: application to adaptor protein complexes in cell signaling. *Protein Sci.*, 16(1):30–42, Jan 2007.

[127] Y. Huang, L. Ji, Q. Huang, D. G. Vassylyev, X. Chen, and J. B. Ma. Structural insights into mechanisms of the small RNA methyltransferase HEN1. *Nature*, 461(7265):823–827, Oct 2009.

[128] W. Humphrey, A. Dalke, and K. Schulten. VMD: visual molecular dynamics. *J Mol Graph*, 14(1):33–38, Feb 1996.

[129] C. M. Hurvich and C.-L Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, . 1989.

[130] Keeler J. *Understanding NMR Spectroscopy (2nd ed.)*. John Wiley and Sons, Ltd, 2010.

[131] B. H. Johnson and M. H. Hecht. Recombinant proteins can be isolated from E. coli cells by repeated cycles of freezing and thawing. *Biotechnology (N.Y.)*, 12(13):1357–1360, Dec 1994.

[132] A.R. Jones. Light scattering for particle characterization. *Progress in Energy and Combustion Science*, 25:1–53, Feb 1999.

[133] D. T. Jones. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, 292(2):195–202, Sep 1999.

[134] R. G. Jones and et al. Analysis of results obtained on biological substances using isothermal titration calorimetry (IUPAC Technical Report). *Pure Appl. Chem.*, 79(10):1801–1829, 2007.

[135] T. Katoh and T. Suzuki. Specific residues at every third position of siRNA shape its efficient RNAi activity. *Nucleic Acids Res.*, 35(4):e27, 2007.

[136] Y. Kawahara, M. Megraw, E. Kreider, H. Iizasa, L. Valente, A. G. Hatzigeorgiou, and K. Nishikura. Frequency and fate of microRNA editing in human brain. *Nucleic Acids Res.*, 36(16):5270–5280, Sep 2008.

[137] Y. Kawahara, B. Zinshteyn, T. P. Chendrimada, R. Shiekhattar, and K. Nishikura. RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. *EMBO Rep.*, 8(8):763–769, Aug 2007.

[138] Y. Kawahara, B. Zinshteyn, P. Sethupathy, H. Iizasa, A. G. Hatzigeorgiou, and K. Nishikura. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science*, 315(5815):1137–1140, Feb 2007.

[139] L. E. Kay, D. A. Torchia, and A. Bax. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, 28(23):8972–8979, Nov 1989.

[140] S. Keller, C. Vargas, H. Zhao, G. Piszczek, C. A. Brautigam, and P. Schuck. High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal. Chem.*, 84(11):5066–5073, Jun 2012.

[141] V. N. Kim, J. Han, and M. C. Siomi. Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.*, 10(2):126–139, Feb 2009.

[142] Y. K. Kim and V. N. Kim. Processing of intronic microRNAs. *EMBO J.*, 26(3):775–783, Feb 2007.

[143] C. Klattenhoff and W. Theurkauf. Biogenesis and germline functions of piRNAs. *Development*, 135(1):3–9, Jan 2008.

[144] I. R. Kleckner and M. P. Foster. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta*, 1814(8):942–968, Aug 2011.

[145] P. Kletting and G. Glatting. Model selection for time-activity curves: the corrected Akaike information criterion and the F-test. *Z Med Phys*, 19(3):200–206, 2009.

[146] K. Kloiber, R. Spitzer, S. Grutsch, C. Kreutz, and M. Tollinger. Longitudinal exchange: an alternative strategy towards quantification of dynamics parameters in ZZ exchange spectroscopy. *J. Biomol. NMR*, 51(1-2):123–129, Sep 2011.

[147] H. R. Koh, M. A. Kidwell, K. Ragunathan, J. A. Doudna, and S. Myong. ATP-independent diffusion of double-stranded RNA binding proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 110(1):151–156, Jan 2013.

[148] K. H. Kok, M. H. Ng, Y. P. Ching, and D. Y. Jin. Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA. *J. Biol. Chem.*, 282(24):17649–17657, Jun 2007.

[149] Y. Kurihara, Y. Takashi, and Y. Watanabe. The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA*, 12(2):206–212, Feb 2006.

[150] Y. Kurihara and Y. Watanabe. Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl. Acad. Sci. U.S.A.*, 101(34):12753–12758, Aug 2004.

[151] M. Kurz, N. P. Cowieson, G. Robin, D. A. Hume, J. L. Martin, B. Kobe, and P. Listwan. Incorporating a TEV cleavage site reduces the solubility of nine recombinant mouse proteins. *Protein Expr. Purif.*, 50(1):68–73, Nov 2006.

[152] E. Ladewig, K. Okamura, A. S. Flynt, J. O. Westholm, and E. C. Lai. Discovery of hundreds of mirtrons in mouse and human small RNA data. *Genome Res.*, 22(9):1634–1645, Sep 2012.

[153] U. K. Laemmli. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259):680–685, Aug 1970.

[154] G. Laraki, G. Clerzius, A. Daher, C. Melendez-Pena, S. Daniels, and A. Gatignol. Interactions between the double-stranded RNA-binding proteins TRBP and PACT define the Medipal domain that mediates protein-protein interactions. *RNA Biol*, 5(2):92–103, 2008.

[155] P. W. Lau, K. Z. Guiley, N. De, C. S. Potter, B. Carragher, and I. J. MacRae. The molecular architecture of human Dicer. *Nat. Struct. Mol. Biol.*, 19(4):436–440, Apr 2012.

[156] P. W. Lau, C. S. Potter, B. Carragher, and I. J. MacRae. Structure of the human Dicer-TRBP complex by electron microscopy. *Structure*, 17(10):1326–1332, Oct 2009.

[157] T.M. Laue, B.D. Shah, T.M. Ridgeway, and S.L. Pelletier. *Computer-Aided Interpretation of Analytical Sedimentation Data For Proteins*. Royal Society of Chemistry; Cambridge, U.K, 1992.

[158] A.L. Lavoisier and P.S. de Laplace. *Mémoire sur la chaleur*. Academie Royale Scientifique, 1784.

[159] M. le Maire, A. Viel, and J. V. Moller. Size exclusion chromatography and universal calibration of gel columns. *Anal. Biochem.*, 177(1):50–56, Feb 1989.

[160] J. Lebowitz, M. S. Lewis, and P. Schuck. Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci.*, 11(9):2067–2079, Sep 2002.

[161] J. Lebowitz, M. S. Lewis, and P. Schuck. Back to the future: A rebuttal to Henryk Eisenberg. *Protein Sci.*, 12(11):2649–2650, Nov 2003.

[162] H. Y. Lee and J. A. Doudna. TRBP alters human precursor microRNA processing in vitro. *RNA*, 18(11):2012–2019, Nov 2012.

[163] L. W. Lee, S. Zhang, A. Etheridge, L. Ma, D. Martin, D. Galas, and K. Wang. Complexity of the microRNA repertoire revealed by next-generation sequencing. *RNA*, 16(11):2170–2180, Nov 2010.

[164] R. C. Lee, R. L. Feinbaum, and V. Ambros. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75(5):843–854, Dec 1993.

[165] S. K. Lee and G. A. Calin. Non-coding RNAs and cancer: new paradigms in oncology. *Discov Med*, 11(58):245–254, Mar 2011.

[166] Y. Lee, C. Ahn, J. Han, H. Choi, J. Kim, J. Yim, J. Lee, P. Provost, O. Radmark, S. Kim, and V. N. Kim. The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 425(6956):415–419, Sep 2003.

[167] Y. Lee, I. Hur, S. Y. Park, Y. K. Kim, M. R. Suh, and V. N. Kim. The role of PACT in the RNA silencing pathway. *EMBO J.*, 25(3):522–532, Feb 2006.

[168] Y. S. Lee, K. Nakahara, J. W. Pham, K. Kim, Z. He, E. J. Sontheimer, and R. W. Carthew. Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell*, 117(1):69–81, Apr 2004.

[169] Kristian Hovde Liland and Bjorn-Helge Mevik. *baseline: Baseline Correction of Spectra*, 2012. R package version 1.1-0.

[170] d. o. H. Lim, J. Kim, S. Kim, R. W. Carthew, and Y. S. Lee. Functional analysis of dicer-2 missense mutations in the siRNA pathway of Drosophila. *Biochem. Biophys. Res. Commun.*, 371(3):525–530, Jul 2008.

[171] W.A. Lindner and Brand J. M. A Coherent Approach to the Svedberg Equation. *Biochem. Education*, 15(2):71–72, 1987.

[172] Q. Liu, T. A. Rand, S. Kalidas, F. Du, H. E. Kim, D. P. Smith, and X. Wang. R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. *Science*, 301(5641):1921–1925, Sep 2003.

[173] X. Liu, F. Jiang, S. Kalidas, D. Smith, and Q. Liu. Dicer-2 and R2D2 coordinately bind siRNA to promote assembly of the siRISC complexes. *RNA*, 12(8):1514–1520, Aug 2006.

[174] X. Liu, J. K. Park, F. Jiang, Y. Liu, D. McKearin, and Q. Liu. Dicer-1, but not Loquacious, is critical for assembly of miRNA-induced silencing complexes. *RNA*, 13(12):2324–2329, Dec 2007.

[175] Y. Liu and C. E. Samuel. Mechanism of interferon action: functionally distinct RNA-binding and catalytic domains in the interferon-inducible, double-stranded RNA-specific adenosine deaminase. *J. Virol.*, 70(3):1961–1968, Mar 1996.

[176] D. Lobbes, G. Rallapalli, D. D. Schmidt, C. Martin, and J. Clarke. SERRATE: a new player on the plant microRNA scene. *EMBO Rep.*, 7(10):1052–1058, Oct 2006.

[177] T. M. Ludden, S. L. Beal, and L. B. Sheiner. Comparison of the Akaike Information Criterion, the Schwarz criterion and the F test as guides to model selection. *J Pharmacokinet Biopharm*, 22(5):431–445, Oct 1994.

[178] E. Ma, I. J. MacRae, J. F. Kirsch, and J. A. Doudna. Autoinhibition of human dicer by its internal helicase domain. *J. Mol. Biol.*, 380(1):237–243, Jun 2008.

[179] E. Ma, K. Zhou, M. A. Kidwell, and J. A. Doudna. Coordinated activities of human dicer domains in regulatory RNA processing. *J. Mol. Biol.*, 422(4):466–476, Sep 2012.

[180] J. B. Ma, K. Ye, and D. J. Patel. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature*, 429(6989):318–322, May 2004.

[181] S. Machida, H. Y. Chen, and Y. Adam Yuan. Molecular insights into miRNA processing by Arabidopsis thaliana SERRATE. *Nucleic Acids Res.*, 39(17):7828–7836, Sep 2011.

[182] I. J. MacRae, E. Ma, M. Zhou, C. V. Robinson, and J. A. Doudna. In vitro reconstitution of the human RISC-loading complex. *Proc. Natl. Acad. Sci. U.S.A.*, 105(2):512–517, Jan 2008.

[183] I. J. MacRae, K. Zhou, and J. A. Doudna. Structural determinants of RNA recognition and cleavage by Dicer. *Nat. Struct. Mol. Biol.*, 14(10):934–940, Oct 2007.

[184] S. Macura, Y. Huang, D. Suter, and R. R. Ernst. Two-dimensional chemical exchange and cross-relaxation spectroscopy of coupled nuclear spins. *J. Magn. Reson.*, 43:259–281, 1981.

[185] S. Macura, W. M. Westler, and J. L. Markley. Two-dimensional exchange spectroscopy of proteins. *Meth. Enzymol.*, 239:106–144, 1994.

[186] E. Maniataki and Z. Mourelatos. A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev.*, 19(24):2979–2990, Dec 2005.

[187] J. T. Marques, K. Kim, P. H. Wu, T. M. Alleyne, N. Jafari, and R. W. Carthew. Loqs and R2D2 act sequentially in the siRNA pathway in Drosophila. *Nat. Struct. Mol. Biol.*, 17(1):24–30, Jan 2010.

[188] G. Masliah, P. Barraud, and F. H. Allain. RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell. Mol. Life Sci.*, Aug 2012.

[189] A. G. Matera, R. M. Terns, and M. P. Terns. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat. Rev. Mol. Cell Biol.*, 8(3):209–220, Mar 2007.

[190] H. M. McConnel. Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.*, 28:430–431, 1958.

[191] L. J. McGuffin, K. Bryson, and D. T. Jones. The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4):404–405, Apr 2000.

[192] S. Melo, A. Villanueva, C. Moutinho, V. Davalos, R. Spizzo, C. Ivan, S. Rossi, F. Setien, O. Casanovas, L. Simo-Riudalbas, J. Carmona, J. Carrere, A. Vidal, A. Aytes, S. Puertas, S. Ropero, R. Kalluri, C. M. Croce, G. A. Calin, and M. Esteller. Small molecule enoxacin is a cancer-specific growth inhibitor that acts by enhancing TAR RNA-binding protein 2-mediated microRNA processing. *Proc. Natl. Acad. Sci. U.S.A.*, 108(11):4394–4399, Mar 2011.

[193] S. A. Melo, S. Ropero, C. Moutinho, L. A. Aaltonen, H. Yamamoto, G. A. Calin, S. Rossi, A. F. Fernandez, F. Carneiro, C. Oliveira, B. Ferreira, C. G. Liu, A. Villanueva, G. Capella, S. Schwartz, R. Shiekhattar, and M. Esteller. A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. *Nat. Genet.*, 41(3):365–370, Mar 2009.

[194] Levitt M.H. *spin dynamics (2nd ed.)*. John Wiley and Sons, Ltd, 2008.

[195] L. A. Mirny and E. I. Shakhnovich. Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. *J. Mol. Biol.*, 291(1):177–196, Aug 1999.

[196] E. A. Miska, E. Alvarez-Saavedra, A. L. Abbott, N. C. Lau, A. B. Hellman, S. M. McGonagle, D. P. Bartel, V. R. Ambros, and H. R. Horvitz. Most Caenorhabditis elegans microRNAs are individually not essential for development or viability. *PLoS Genet.*, 3(12):e215, Dec 2007.

[197] K. Miyoshi, T. Miyoshi, J. V. Hartig, H. Siomi, and M. C. Siomi. Molecular mechanisms that funnel RNA precursors into endogenous small-interfering RNA and microRNA biogenesis pathways in Drosophila. *RNA*, 16(3):506–515, Mar 2010.

[198] K. Miyoshi, T. Miyoshi, and H. Siomi. Many ways to generate microRNA-like small RNAs: non-canonical pathways for microRNA production. *Mol. Genet. Genomics*, 284(2):95–103, Aug 2010.

[199] L. S. Mizoue and J. Tellinghuisen. The role of backlash in the first injection anomaly in isothermal titration calorimetry. *Anal. Biochem.*, 326(1):125–127, Mar 2004.

[200] Y. K. Mok, E. L. Elisseeva, A. R. Davidson, and J. D. Forman-Kay. Dramatic stabilization of an SH3 domain by a single substitution: roles of the folded and unfolded states. *J. Mol. Biol.*, 307(3):913–928, Mar 2001.

[201] G. A. Mueller, M. T. Miller, E. F. Derose, M. Ghosh, R. E. London, and T. M. Hall. Solution structure of the Drosha double-stranded RNA-binding domain. *Silence*, 1(1):2, 2010.

[202] F. A. Mulder, D. Schipper, R. Bott, and R. Boelens. Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins. *J. Mol. Biol.*, 292(1):111–123, Sep 1999.

[203] S. Nanduri, B. W. Carpick, Y. Yang, B. R. Williams, and J. Qin. Structure of the double-stranded RNA-binding domain of the protein kinase PKR reveals the molecular basis of its dsRNA-mediated activation. *EMBO J.*, 17(18):5458–5465, Sep 1998.

[204] S. Nanduri, F. Rahman, B. R. Williams, and J. Qin. A dynamically tuned double-stranded RNA binding mechanism for the activation of antiviral kinase PKR. *EMBO J.*, 19(20):5567–5574, Oct 2000.

[205] M. Negrini, M. S. Nicoloso, and G. A. Calin. MicroRNAs and cancer–new paradigms in molecular oncology. *Curr. Opin. Cell Biol.*, 21(3):470–479, Jun 2009.

[206] V. Neuhoff, R. Stamm, I. Pardowitz, N. Arold, W. Ehrhardt, and D. Taube. Essential problems in quantification of proteins following colloidal staining with coomassie brilliant blue dyes in polyacrylamide gels, and their solution. *Electrophoresis*, 11(2):101–117, Feb 1990.

[207] C. L. Noland, E. Ma, and J. A. Doudna. siRNA repositioning for guide strand selection by human Dicer complexes. *Mol. Cell*, 43(1):110–121, Jul 2011.

[208] M. A. O'Connell and L. P. Keegan. Drosha versus ADAR: wrangling over pri-miRNA. *Nat. Struct. Mol. Biol.*, 13(1):3–4, Jan 2006.

[209] M. Ohman. A-to-I editing challenger or ally to the microRNA process. *Biochimie*, 89(10):1171–1176, Oct 2007.

[210] C. Okada, E. Yamashita, S. J. Lee, S. Shibata, J. Katahira, A. Nakagawa, Y. Yoneda, and T. Tsukihara. A high-resolution structure of the pre-microRNA nuclear export machinery. *Science*, 326(5957):1275–1279, Nov 2009.

[211] K. Okamura, J. W. Hagen, H. Duan, D. M. Tyler, and E. C. Lai. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. *Cell*, 130(1):89–100, Jul 2007.

[212] K. Okamura, N. Robine, Y. Liu, Q. Liu, and E. C. Lai. R2D2 organizes small regulatory RNA pathways in Drosophila. *Mol. Cell. Biol.*, 31(4):884–896, Feb 2011.

[213] J. E. Park, I. Heo, Y. Tian, D. K. Simanshu, H. Chang, D. Jee, D. J. Patel, and V. N. Kim. Dicer recognizes the 5′ end of RNA for efficient and accurate processing. *Nature*, 475(7355):201–205, Jul 2011.

[214] G. S. Parker, D. M. Eckert, and B. L. Bass. RDE-4 preferentially binds long dsRNA and its dimerization is necessary for cleavage of dsRNA to siRNA. *RNA*, 12(5):807–818, May 2006.

[215] Z. Paroo, X. Ye, S. Chen, and Q. Liu. Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling. *Cell*, 139(1):112–122, Oct 2009.

[216] C. V. Patel, I. Handy, T. Goldsmith, and R. C. Patel. PACT, a stress-modulated cellular activator of interferon-induced double-stranded RNA-activated protein kinase, PKR. *J. Biol. Chem.*, 275(48):37993–37998, Dec 2000.

[217] R. C. Patel and G. C. Sen. PACT, a protein activator of the interferon-induced protein kinase, PKR. *EMBO J.*, 17(15):4379–4390, Aug 1998.

[218] R. C. Patel, P. Stanton, N. M. McMillan, B. R. Williams, and G. C. Sen. The interferon-inducible double-stranded RNA-activated protein kinase self-associates in vitro and in vivo. *Proc. Natl. Acad. Sci. U.S.A.*, 92(18):8283–8287, Aug 1995.

[219] C.L. Perrin and T.J. Dwyer. Application of two-dimensional NMR to kinetics of chemical exchange. *Chem. Rev.*, 90:935–967, Sep 1990.

[220] G. A. Peters, R. Hartmann, J. Qin, and G. C. Sen. Modular structure of PACT: distinct domains for binding and activating PKR. *Mol. Cell. Biol.*, 21(6):1908–1920, Mar 2001.

[221] R. Pitt-Rivers and F. S. Impiombato. The binding of sodium dodecyl sulphate to various proteins. *Biochem. J.*, 109(5):825–830, Oct 1968.

[222] J. R. Platt. Strong Inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science*, 146(3642):347–353, Oct 1964.

[223] S. Podzimek. *Light Scattering, Size Exclusion Chromatography and Asymmetric Flow Field*. Wiley, 2011.

[224] G. M. Poon. Explicit formulation of titration models for isothermal titration calorimetry. *Anal. Biochem.*, 400(2):229–236, May 2010.

[225] M. Potschka. Universal calibration of gel permeation chromatography and determination of molecular shape in solution. *Anal. Biochem.*, 162(1):47–64, Apr 1987.

[226] P. Provost, D. Dishart, J. Doucet, D. Frendewey, B. Samuelsson, and O. Radmark. Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J.*, 21(21):5864–5874, Nov 2002.

[227] H. Qin, F. Chen, X. Huan, S. Machida, J. Song, and Y. A. Yuan. Structure of the Arabidopsis thaliana DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction. *RNA*, 16(3):474–481, Mar 2010.

[228] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010.

[229] A. Ramos, S. Grunert, J. Adams, D. R. Micklem, M. R. Proctor, S. Freund, M. Bycroft, D. St Johnston, and G. Varani. RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J.*, 19(5):997–1009, Mar 2000.

[230] T. A. Rand, S. Petersen, F. Du, and X. Wang. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell*, 123(4):621–629, Nov 2005.

[231] R. M. Rasia, J. Mateos, N. G. Bologna, P. Burdisso, L. Imbert, J. F. Palatnik, and J. Boisbouvier. Structure and RNA interactions of the plant MicroRNA processing-associated protein HYL1. *Biochemistry*, 49(38):8237–8239, Sep 2010.

[232] G. B. Robb and T. M. Rana. RNA helicase A interacts with RISC in human cells and functions in RISC loading. *Mol. Cell*, 26(4):523–537, May 2007.

[233] J. J. Rossi. Mammalian Dicer finds a partner. *EMBO Rep.*, 6(10):927–929, Oct 2005.

[234] B. M. Roth and M. Hennig. Backbone (1)H (N), (13)C, and (15)N resonance assignments of the tandem RNA-binding domains of human DGCR8. *Biomol NMR Assign*, Jul 2012.

[235] J. G. Ruby, C. Jan, C. Player, M. J. Axtell, W. Lee, C. Nusbaum, H. Ge, and D. P. Bartel. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell*, 127(6):1193–1207, Dec 2006.

[236] J. G. Ruby, C. H. Jan, and D. P. Bartel. Intronic microRNA precursors that bypass Drosha processing. *Nature*, 448(7149):83–86, Jul 2007.

[237] J. M. Ryter and S. C. Schultz. Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J.*, 17(24):7505–7513, Dec 1998.

[238] K. Sakurai, M. Amarzguioui, D. H. Kim, J. Alluin, B. Heale, M. S. Song, A. Gatignol, M. A. Behlke, and J. J. Rossi. A role for human Dicer in pre-RISC loading of siRNAs. *Nucleic Acids Res.*, 39(4):1510–1525, Mar 2011.

[239] T. Sasaki, A. Shiohama, S. Minoshima, and N. Shimizu. Identification of eight members of the Argonaute family in the human genome small star, filled. *Genomics*, 82(3):323–330, Sep 2003.

[240] L. R. Saunders and G. N. Barber. The dsRNA binding protein family: critical roles, diverse cellular functions. *FASEB J.*, 17(9):961–983, Jun 2003.

[241] P. Schuck. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.*, 78(3):1606–1619, Mar 2000.

[242] P. Schuck. Diffusion of the reaction boundary of rapidly interacting macromolecules in sedimentation velocity. *Biophys. J.*, 98(11):2741–2751, Jun 2010.

[243] P. Schuck. Sedimentation patterns of rapidly reversible protein interactions. *Biophys. J.*, 98(9):2005–2013, May 2010.

[244] F.P. Schwarz, T. Reinisch, HJ Hinz, and A. Surolia. Analysis of results obtained on biological substances using isothermal titration calorimetry (IUPAC Technical Report). *Pure Appl. Chem.*, 80(9):2025–2040, 2008.

[245] S. A. Shabalina and E. V. Koonin. Origins and evolution of eukaryotic RNA interference. *Trends Ecol. Evol. (Amst.)*, 23(10):578–587, Oct 2008.

[246] C. E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27, 1948.

[247] Y. Shen, F. Delaglio, G. Cornilescu, and A. Bax. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, 44:213–223, Aug 2009.

[248] M. Singh, D. Castillo, C. V. Patel, and R. C. Patel. Stress-induced phosphorylation of PACT reduces its interaction with TRBP and leads to PKR activation. *Biochemistry*, 50(21):4550–4560, May 2011.

[249] L. Sinkkonen, T. Hugenschmidt, W. Filipowicz, and P. Svoboda. Dicer is associated with ribosomal DNA chromatin in mammalian cells. *PLoS ONE*, 5(8):e12175, 2010.

[250] A. Sivashanmugam, V. Murray, C. Cui, Y. Zhang, J. Wang, and Q. Li. Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. *Protein Sci.*, 18(5):936–948, May 2009.

[251] N. R. Smalheiser. The search for endogenous siRNAs in the mammalian brain. *Exp. Neurol.*, 235(2):455–463, Jun 2012.

[252] S. Y. Sohn, W. J. Bae, J. J. Kim, K. H. Yeom, V. N. Kim, and Y. Cho. Crystal structure of human DGCR8 core. *Nat. Struct. Mol. Biol.*, 14(9):847–853, Sep 2007.

[253] R. Song, G. W. Hennig, Q. Wu, C. Jose, H. Zheng, and W. Yan. Male germ cells express abundant endogenous siRNAs. *Proc. Natl. Acad. Sci. U.S.A.*, 108(32):13159–13164, Aug 2011.

[254] D. St Johnston, N. H. Brown, J. G. Gall, and M. Jantsch. A conserved double-stranded RNA-binding domain. *Proc. Natl. Acad. Sci. U.S.A.*, 89(22):10979–10983, Nov 1992.

[255] P. B. Stathopulos, G. A. Scholz, Y. M. Hwang, J. A. Rumfeldt, J. R. Lepock, and E. M. Meiering. Sonication of proteins causes formation of aggregates that resemble amyloid. *Protein Sci.*, 13(11):3017–3027, Nov 2004.

[256] R. Stefl, F. C. Oberstrass, J. L. Hood, M. Jourdan, M. Zimmermann, L. Skrisovska, C. Maris, L. Peng, C. Hofr, R. B. Emeson, and F. H. Allain. The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell*, 143(2):225–237, Oct 2010.

[257] R. Stefl, M. Xu, L. Skrisovska, R. B. Emeson, and F. H. Allain. Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs. *Structure*, 14(2):345–355, Feb 2006.

[258] X. Sun, J. M. Li, and R. M. Wartell. Conversion of stable RNA hairpin to a metastable dimer in frozen solution. *RNA*, 13(12):2277–2286, Dec 2007.

[259] I. D. Swan. The inhibition of hen egg-white lysozyme by imidazole and indole derivatives. *J. Mol. Biol.*, 65(1):59–62, Mar 1972.

[260] H. Tabara, E. Yigit, H. Siomi, and C. C. Mello. The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans. *Cell*, 109(7):861–871, Jun 2002.

[261] Y. Tagami, H. Motose, and Y. Watanabe. A dominant mutation in DCL1 suppresses the hyl1 mutant phenotype by promoting the processing of miRNA. *RNA*, 15(3):450–458, Mar 2009.

[262] S. N. Taylor, C. L. Wagner, and B. W. Hollis. Maternal or Neonatal Vitamin D Supplementation During Lactation: What Is the Better Option? *Annu. Rev. Nutr.*, Jun 2006.

[263] S. S. Taylor, N. M. Haste, and G. Ghosh. PKR and eIF2alpha: integration of kinase dimerization, activation, and substrate docking. *Cell*, 122(6):823–825, Sep 2005.

[264] B. Tian, P. C. Bevilacqua, A. Diegelman-Parente, and M. B. Mathews. The double-stranded-RNA-binding motif: interference and much more. *Nat. Rev. Mol. Cell Biol.*, 5(12):1013–1023, Dec 2004.

[265] B. Tian and M. B. Mathews. Functional characterization of and cooperation between the double-stranded RNA-binding motifs of the protein kinase PKR. *J. Biol. Chem.*, 276(13):9936–9944, Mar 2001.

[266] Y. Tomari, T. Du, and P. D. Zamore. Sorting of Drosophila small silencing RNAs. *Cell*, 130(2):299–308, Jul 2007.

[267] Y. Tomari, C. Matranga, B. Haley, N. Martinez, and P. D. Zamore. A protein sensor for siRNA asymmetry. *Science*, 306(5700):1377–1380, Nov 2004.

[268] Y. Tomari and P. D. Zamore. Perspective: machines for RNAi. *Genes Dev.*, 19(5):517–529, Mar 2005.

[269] P. Tompa and A. Fersht. *Structure and Function of Intrinsically Disordered Proteins*. Chapman and Hall/CRC, 2009.

[270] J. W. Ucci, Y. Kobayashi, G. Choi, A. T. Alexandrescu, and J. L. Cole. Mechanism of interaction of the double-stranded RNA (dsRNA) binding domain of protein kinase R with short dsRNA sequences. *Biochemistry*, 46(1):55–65, Jan 2007.

[271] V. N. Uversky. Size-exclusion chromatography in structural analysis of intrinsically disordered proteins. *Methods Mol. Biol.*, 896:179–194, 2012.

[272] J. A. van Dijk and J. A. Smit. Size-exclusion chromatography–multiangle laser light scattering analysis of beta-lactoglobulin and bovine serum albumin in aqueous solution with added salt. *J Chromatogr A*, 867(1-2):105–112, Jan 2000.

[273] A. Vera, N. Gonzalez-Montalban, A. Aris, and A. Villaverde. The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnol. Bioeng.*, 96(6):1101–1106, Apr 2007.

[274] O. Voinnet. Non-cell autonomous RNA silencing. *FEBS Lett.*, 579(26):5858–5871, Oct 2005.

[275] Z. Wang, E. Hartman, K. Roy, G. Chanfreau, and J. Feigon. Structure of a yeast RNase III dsRBD complex with a noncanonical RNA substrate provides new insights into binding specificity of dsRBDs. *Structure*, 19(7):999–1010, Jul 2011.

[276] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, and D. T. Jones. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.*, 337(3):635–645, Mar 2004.

[277] N. C. Welker, T. S. Maity, X. Ye, P. J. Aruscavage, A. A. Krauchuk, Q. Liu, and B. L. Bass. Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode. *Mol. Cell*, 41(5):589–599, Mar 2011.

[278] J. Wen, T. Arakawa, and J. S. Philo. Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. *Anal. Biochem.*, 240(2):155–166, Sep 1996.

[279] J. O. Westholm and E. C. Lai. Mirtrons: microRNA biogenesis via splicing. *Biochimie*, 93(11):1897–1904, Nov 2011.

[280] T. Wiseman, S. Williston, J. F. Brandts, and L. N. Lin. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.*, 179(1):131–137, May 1989.

[281] D. S. Wishart, B. D. Sykes, and F. M. Richards. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.*, 222(2):311–333, Nov 1991.

[282] C. J. Wong, K. Launer-Felty, and J. L. Cole. Analysis of PKR-RNA interactions by sedimentation velocity. *Meth. Enzymol.*, 488:59–79, 2011.

[283] C. Wostenberg, J. W. Lary, D. Sahu, R. Acevedo, K. A. Quarles, J. L. Cole, and S. A. Showalter. The role of human Dicer-dsRBD in processing small regulatory RNAs. *PLoS ONE*, 7(12):e51829, 2012.

[284] C. Wostenberg, K. A. Quarles, and S. A. Showalter. Dynamic origins of differential RNA binding function in two dsRBDs from the miRNA microprocessor complex. *Biochemistry*, 49(50):10728–10736, Dec 2010.

[285] H. Wu, A. Henras, G. Chanfreau, and J. Feigon. Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. *Proc. Natl. Acad. Sci. U.S.A.*, 101(22):8307–8312, Jun 2004.

[286] P.J. Wyatt. Light-scattering and the absolute characterization of macromolecules. *Anal Chim Acta*, 272:1–40, 1993.

[287] J. Xia, C. E. Joyce, A. M. Bowcock, and W. Zhang. Noncanonical microRNAs and endogenous siRNAs in normal and psoriatic human skin. *Hum. Mol. Genet.*, 22(4):737–748, Feb 2013.

[288] T. Xie, J. Penelle, and M. Verraver. Experimental investigation on the reliability of routine SEC–MALLS for the determination of absolute molecular weights in the oligomeric range. *Polymer*, 43:3973–3977, 2002.

[289] Z. Xie, L. K. Johansen, A. M. Gustafson, K. D. Kasschau, A. D. Lellis, D. Zilberman, S. E. Jacobsen, and J. C. Carrington. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.*, 2(5):E104, May 2004.

[290] S. Yamashita, T. Nagata, M. Kawazoe, C. Takemoto, T. Kigawa, P. Guntert, N. Kobayashi, T. Terada, M. Shirouzu, M. Wakiyama, Y. Muto, and S. Yokoyama. Structures of the first and second double-stranded RNA-binding domains of human TAR RNA-binding protein. *Protein Sci.*, 20(1):118–130, Jan 2011.

[291] J. S. Yang and E. C. Lai. Dicer-independent, Ago2-mediated microRNA biogenesis in vertebrates. *Cell Cycle*, 9(22):4455–4460, Nov 2010.

[292] J. S. Yang and E. C. Lai. Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. *Mol. Cell*, 43(6):892–903, Sep 2011.

[293] J. S. Yang, T. Maurin, N. Robine, K. D. Rasmussen, K. L. Jeffrey, R. Chandwani, E. P. Papapetrou, M. Sadelain, D. O'Carroll, and E. C. Lai. Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. *Proc. Natl. Acad. Sci. U.S.A.*, 107(34):15163–15168, Aug 2010.

[294] S. W. Yang, H. Y. Chen, J. Yang, S. Machida, N. H. Chua, and Y. A. Yuan. Structure of Arabidopsis HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. *Structure*, 18(5):594–605, May 2010.

[295] W. Yang, T. P. Chendrimada, Q. Wang, M. Higuchi, P. H. Seeburg, R. Shiekhattar, and K. Nishikura. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat. Struct. Mol. Biol.*, 13(1):13–21, Jan 2006.

[296] X. Ye, Z. Paroo, and Q. Liu. Functional anatomy of the Drosophila microRNA-generating enzyme. *J. Biol. Chem.*, 282(39):28373–28378, Sep 2007.

[297] P. D. Zamore and B. Haley. Ribo-gnome: the big world of small RNAs. *Science*, 309(5740):1519–1524, Sep 2005.

[298] H. Zhang, F. A. Kolb, V. Brondani, E. Billy, and W. Filipowicz. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.*, 21(21):5875–5885, Nov 2002.

[299] H. Zhang, F. A. Kolb, L. Jaskiewicz, E. Westhof, and W. Filipowicz. Single processing center models for human Dicer and bacterial RNase III. *Cell*, 118(1):57–68, Jul 2004.

[300] L. Zhang, D. Hou, X. Chen, D. Li, L. Zhu, Y. Zhang, J. Li, Z. Bian, X. Liang, X. Cai, Y. Yin, C. Wang, T. Zhang, D. Zhu, D. Zhang, J. Xu, Q. Chen, Y. Ba, J. Liu, Q. Wang, J. Chen, J. Wang, M. Wang, Q. Zhang, J. Zhang, K. Zen, and C. Y. Zhang. Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. *Cell Res.*, 22(1):107–126, Jan 2012.

[301] O. Zhang and J. D. Forman-Kay. Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer. *Biochemistry*, 34(20):6784–6794, May 1995.

[302] O. Zhang, L. E. Kay, J. P. Olivier, and J. D. Forman-Kay. Backbone 1H and 15N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR*, 4(6):845–858, Nov 1994.

[303] H. Zhao, A. Balbo, P. H. Brown, and P. Schuck. The boundary structure in the analysis of reversibly interacting systems by sedimentation velocity. *Methods*, 54(1):16–30, May 2011.

[304] H. Zhao, P. H. Brown, and P. Schuck. On the distribution of protein refractive index increments. *Biophys. J.*, 100(9):2309–2317, May 2011.

[305] X. Zheng and P. C. Bevilacqua. Straightening of bulged RNA by the double-stranded RNA-binding domain from the protein kinase PKR. *Proc. Natl. Acad. Sci. U.S.A.*, 97(26):14162–14167, Dec 2000.

[306] R. Zhou, B. Czech, J. Brennecke, R. Sachidanandam, J. A. Wohlschlegel, N. Perrimon, and G. J. Hannon. Processing of Drosophila endo-siRNAs depends on a specific Loquacious isoform. *RNA*, 15(10):1886–1895, Oct 2009.

[307] B.H. Zimm. Molecular Theory of the Scattering of Light in Fluids. *J. Chem. Phys.*, 141, 1945.

[308] M. Zolkiewski and A. Ginsburg. Thermodynamic effects of active-site ligands on the reversible, partial unfolding of dodecameric glutamine synthetase from Escherichia coli: calorimetric studies. *Biochemistry*, 31(48):11991–12000, Dec 1992.

[309] M. Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31(13):3406–3415, Jul 2003.

Résumé en français Les microARNs (miARNs) sont une classe de petits ARNs qui régulent l'expression des gènes. Lors de la biogenèse des miARNs humains, le précurseur de miARN est reconnu et clivé par le complexe Dicer/TRBP. La thèse caractérise en solution la région d'interaction de TRBP avec les ARNs à l'état libre et à l'état lié par le précurseur du miARN oncogène miR-155 et par son produit après clivage par Dicer/TRBP. Une combinaison de méthodes biophysiques (RMN, ITC, AUC, SEC, MALLS) a été utilisée pour établir les propriétés des interactions. Les analyses montrent notamment que la région d'interaction de TRBP avec les ARNs est composée, à l'état libre ou à l'état lié, de deux domaines de liaison aux ARN doubles brins de conformations indépendantes en solution et dont l'un est le siège d'un échange plié/déplié intégral. La similarité des stœchiométries, affinités et surfaces d'interactions entre TRBP et les miARNs immatures avant et après leur clivage par Dicer suggère de nouveaux rôles pour TRBP

Résumé en anglais MicroRNA (miRNAs) are a class of small RNAs that regulate gene expression. During the biogenesis of human miRNAs, a precursor of miRNA is recognized and cleaved by the complex Dicer/TRBP. The thesis characterizes in solution the interaction region of TRBP with ARNs in the free and bound state by the precursor of the oncogenic microRNA 155 and by its product after cleavage by Dicer/TRBP. A combination of biophysical methods (NMR, ITC, AUC, SEC, MALLS) were used to establish the interaction properties. The analyses show that the RNA binding region of TRBP is composed, in the free and bound state, of two double stranded RNA binding domains of independent conformations in solution, one of them undergoing a full folded/unfolded exchange. The similarity of the stoichiometries, affinities and interaction surfaces between TRBP and miRNA precursors before and after their clivage by Dicer suggests new roles for TRBP.